

Technology transfer by stock exchanges

Takeover battles for technically orientated companies are all the rage, but they will embody an element of shadow-boxing until stock exchanges become technically more expert than they are.

THESE are heady days for the people called investors who sustain the world's stock markets with the commissions they pay to the brokers who buy and sell shares in public companies. On Monday last week, the New York Stock Exchange found itself in delicious turmoil after IBM offered to buy the software manufacturer called Lotus for a cool \$3.3 billion; with all that money on offer more than merely the holders of Lotus shares could expect to benefit. But IBM's adventure is but one of many to have generated excitement in the past year. Microsoft offered roughly the same amount for its much smaller rival Intuit until the US Justice Department announced an anti-trust inquiry; the deal was promptly cancelled. In both cases the supposed victim of the deal boasted a technical innovation that the larger company coveted — networking software called 'Notes' (Lotus) and financial software (Intuit) respectively. Neither company can have thought its technical edge was as valuable as the offers made for it. Intuit is said to have been so unsettled by the experience that it is looking around for another predator whose pockets are as deep as Microsoft's.

A little reflection will confirm that software is not the only technology being bought and sold on stock exchanges. Last year, the British-based international drug company Glaxo bought its smaller rival Wellcome for £10 billion. Now the company called Zeneca, formerly part of the British-based multinational ICI, is variously touted as being "on the block" (meaning that it may profitably be bought) or, alternatively that it maybe compelled by the logic of accountancy to buy some other company. What the investment pundits say is that Zeneca has more innovative drugs in development than it can hope to sell effectively and has therefore two choices; to be sold to a marketing company or to buy a sales force for itself. That, at least, is what the pundits say. At least on the face of things, these are strange ways of bring technical projects to fruition.

Two issues arise, of which the most obvious is that company mergers of this kind are indeed ways of transferring technology from one place to another. That, for example, is the motive underlying the repeated purchase of small biotechnology or software companies by larger companies in their field, especially in the United States. Does it not make sense that such a company should carry through the development of a single idea with the help of investors' funds and then tumble into the arms of a larger competitor capable of exploiting the innovation effectively? The economic benefits are made real more quickly than would

otherwise be the case while the investors get their money back and more. But these purchasers are not always as diligent at exploiting the companies they buy as economic considerations would dictate. And the purchased company may turn out to be less valuable than it appeared because its skilled people learn to do other things. That process is so familiar in the United States that it is simply described by the verb "to walk".

The second issue is that the world's stock exchanges are probably consistently inaccurate in estimating the value of technically based companies. It is not merely that the ultimate providers of the cash for technical mergers are rarely technical people, but that there is no easy way of calculating the economic benefits of a purchase to a larger entity. It is true that stockbrokers and banks increasingly employ technically qualified people as analysts of market trends, but the perceptiveness of such people is often limited by the information they can gather. At the same time, they are compelled by the circumstances in which they work to follow the ethos of their local stock market. Often, strictly short-term calculations of financial advantage dominate. In Britain, for example, the convention seems to be that those who back a novel venture become increasingly impatient if their money has not been returned within three years. After five they become actively discontented. As it happens, the London Stock Exchange is launching next Monday a new market for shares in small companies, unhelpfully called Alternative Investment Markets. It remains to be seen whether that will be more successful than its predecessor, the Unlisted Securities Market, at stimulating the flow of funds to backing innovations.

What will happen to IBM's bid for Lotus is no doubt a matter that IBM can decide for itself; it is a grown-up company with the resources to study the consequences of a merger. (Those who hold IBM's shares are understandably nervous, but that is not an unfamiliar hazard.) The much more serious impediment to adventurous investment is that the same investors may be persuaded to keep their wallets in their pockets. But the days have gone when people could hope to raise funds on the personal reputations of those who found companies and launch them on the stock exchanges. The obvious remedy is that the stock exchanges and their stockbrokers should employ even more technical analysts, but that would not rid them of present constraints. Much better that the ultimate providers of investment cash should know more clearly what they are backing and why



(and in what circumstances) the investment may be worthwhile. That amounts to a further plea for a better public understanding of science — one that illustrates the financial and economic hazards of continued ignorance. □

Honorary authorship

A case of fraud at a London hospital points to the need for a ban on honorary co-authorship in the literature.

THE British, who are given to self-contentment bordering on smugness, are fond of saying that fraud in science does not happen in what Shakespeare called “this scepter’d isle” (chosen as the title of a new 200-part series on BBC radio). After last week, they will no doubt be claiming that part of the explanation is that the penalties for being found out are severe and swift: last week, the British Medical Council “struck off” the national medical register (or delicensed) a physician who had falsely claimed to have transplanted an ectopic pregnancy from its aberrant position to the uterus of the woman concerned, and that the fetus had developed normally and had been delivered successfully (see *Nature* 372, 390; 1994). The evidence given last week appears to have been convincing, but Dr Malcolm Pearce says he will appeal against the council’s findings and decision.

Certainly the British case, which first came to light six months ago, has been handled more swiftly than the US case of Dr Thereza Imanishi-Kari. She has been accused by the US National Institutes of Health of fabricating data in a paper published nine years ago and whose best-known author is Dr David Baltimore. By coincidence, her appeal is being heard this week in Washington, DC, by an administrative tribunal of the Department of Health and Human Services. Imanishi-Kari has been on the research community’s equivalent of death row for much too long. Everybody, for a variety of reasons, will agree on that. But it is also possible that the British case has been too swiftly, or too unreflectively, decided.

The circumstances betoken muddled principles. The offending Pearce paper was published in the *British Journal of Obstetrics and Gynaecology* last August. Pearce boasted two co-authors: a more junior person than he and the head of his laboratory, Professor Geoffrey Chamberlain. Chamberlain was editor-in-chief of the journal in which Pearce’s account of his surgery was published, but he resigned earlier this year. (Honourably, he also then resigned as president of the professional qualifying college to which the journal belongs.) Throughout the affair, Chamberlain has insisted that he knew nothing of the details of the surgical procedure that Pearce had reported, and that he had agreed that his name should be added to the published paper “as a courtesy”.

“There’s the rub” (to quote Shakespeare again). For a long time, and certainly since the investigation by Ned Feder and Walter W. Stewart of the publication habits of the co-authors of Dr John Darsee at the Harvard Medical School and elsewhere (*Nature* 325, 207–214; 1987), the

perils of ‘honorary coauthorship’ have been plain. Persons whose reputations flourish when they are among the authors of respectable research are damaged if the research turns out to be based on lies. Mostly, of course, the research is passable, if dull, and no harm is done. What should happen when it appears to have been concocted? Given prevailing local conventions, it would be too much to ask that Chamberlain should have joined Pearce in the dock, but he has a residuary responsibility: he should publicly acknowledge that he should not have allowed his name to appear among the authors of a paper in which he now claims (in his own defence) to have played no part. How else will others like him learn prudence? □

Masculinity at risk

The discovery that the major metabolite of DDT may damage male reproduction deserves attention.

It would be a cruel irony if the introduction of DDT half a century ago turned out to have been responsible for a decline of male fertility worldwide, but that is one reading of the report by W. R. Kelce and colleagues on page 581 of this issue and the accompanying comment (on page 538) by Dr Richard M Sharpe. What emerges is that the major metabolite of DDT, *p,p'*-DDE, is a demasculinizing agent. As Sharpe points out, because DDT is long-lived in the environment (and still widely used against malarial mosquitoes) there is at least a possibility that its ubiquity in human fat may be responsible for falling sperm counts and other reported abnormalities of male reproduction. So what should be done?

The first need is for a better understanding. To be sure, that will seem a fatuous opinion to those who will regard the new developments as threats to their personal health and who may also fear that reduced fertility may spell the end of the human race. The plain truth is that it would be more alarming if the recent increase in the rate of occurrence of testicular cancer (which Denmark is taking seriously) were linked with *p,p'*-DDE and were the tip of a larger iceberg. But the immediate need, in that context and others, is for a knowledge of whatever relationship there may be between individuals’ content of DDT and the risk that they will develop cancer associated with the reproductive system. That will be a difficult, but not an impossible, undertaking.

Meanwhile, there is a need for a certain calm. Panic will not make DDT disappear from the environment. Moreover, this will not be the first occasion when a presumed link between an environmental contaminant and a risk to health has melted away on close investigation. It cannot but be relevant that the most serious burdens on human health at present are the old ones — resurgent malaria, resurgent tuberculosis. It would be self-defeating (and politically impossible) to push for the outright banning of DDT in, say, mosquito control just in case the presumed effects of its metabolites on male fertility are substantiated. But the case for research to be conducted urgently is overwhelming. □

Physics caught in fight over future plans for US energy department

Washington. A panel of newly elected Republicans in the US House of Representatives has set out plans to disband the Department of Energy (DoE). It proposes appointing a 'closure commission' to rationalize the department's world-renowned network of physics laboratories, and transferring responsibility for nuclear weapons research and development to the Pentagon.

This drastic proposal marks the starting point of a high-stakes game of power politics in Washington, the outcome of which will reshape support for the physical sciences in the United States. Last year, the Department of Energy spent \$6.5 billion on research and development, most of it on nuclear weapons, energy supply, and high-energy and nuclear physics.

There are three possible futures in sight for DoE: disbandment, as proposed by the House panel; the absorption of its research activities into a new Department of Science, as championed by Bob Walker (Republican, Pennsylvania), chair of the House Science Committee; or its retention on a considerably reduced scale, as advocated by the Clinton administration and the energy secretary, Hazel O'Leary.

Disbandment was endorsed at a press

conference last week by Bob Dole (Republican, Kansas), majority leader in the Senate and presidential candidate, Dick Army (Republican, Texas), majority leader in the House, and John Kasich (Republican, Ohio), chairman of the House Budget Committee.

This proposal would place the three weapons laboratories — Los Alamos, Sandia and Lawrence Livermore — in a new, civilian, Defense Nuclear Programs Agency inside the Department of Defense (DoD), along with the DoE's vast nuclear clean-up programme.

The rest of the DoE's 30 or so laboratories would be handed over to an Energy Laboratory Facilities Commission. This would recommend a programme of laboratory closures and consolidations which Congress would have to accept in one package, to avoid endless political wrangling.

"There are far too many laboratories, and far too much 'pork'," said Kasich. "We badly need a laboratory closure commission." Defending the plans to bring nuclear weapons research into DoD, Todd Tiaht (Republican, Kansas), chair of the freshmen's panel, said that the Manhattan Project to build the atomic bomb had

worked under military control.

Kasich warned the freshmen that "we're going to have to fight some of our own people to get this done," and the proposal is likely to meet significant opposition from senators with laboratories in their states, most notably Pete Domenici (Republican, New Mexico), chair of the Senate Budget Committee.

While some continue to make the administrative and ethical case for civilian control of nuclear weapons research, Domenici



O'Leary: faces gulf with laboratories.

will be more concerned about the prospective abilities of Los Alamos and Sandia to retain their budgets inside DoD, at a time when there are no plans for new nuclear weapons. At the same time, environmentalists may react strongly to

the idea that nuclear clean-up should come within the secretive web of the Pentagon.

Thus, despite its heavyweight backing, the freshmen's proposal raises obvious problems, as anticipated by Walker's proposal for a Department of Science (see *Nature* 374, 201; 1995). Having failed to incorporate the proposal in the House Republican budget, Walker is not discouraged: "As we go further down the road [of determining the DoE's future] the concept of a Department of Science will come to the fore," he predicts.

Plans for such a department — which would also incorporate the National Science Foundation, the National Aeronautics and Space Administration (NASA) and several other science agencies — have been criticized by Jack Gibbons, President Clinton's science adviser, and others on the grounds that they would separate science at DoE and NASA from the missions that they are supposed to support.

In response, supporters claim money for science would be safer if it were clearly labelled as such by being placed inside a Department of Science. Even in this Congress, no-one ever has a bad word for science — but it is still cut, the argument goes, because it rests inside unpopular agencies.

The Department of Science solution is certainly more elegant than its critics often admit, but it lacks political support outside the office of the man who would oversee the new department, namely Walker himself.

That leaves a slimmed-down Department of Energy as the most probable alterna- ►

MIT nuclear laboratory feels chill wind

Boston. Shock-waves passed through the Massachusetts Institute of Technology (MIT) last week when it was informed that a congressional subcommittee was planning to cut off funds for the university's Bates Linear Accelerator.

Two days later, the House of Representatives' energy and environment subcommittee voted on a proposal to restore the facility's \$18.6-million budget. But the amendment was defeated in a 12-12 tie.

Four other university-run accelerators — at Yale, Texas A&M, the University of Washington, and the Triangle Universities Nuclear Laboratory in North Carolina — also face closure as a result of proposed Department of Energy (DoE) budget cuts totalling \$31 million (see above).

The subcommittee plan still has a long way to go in Congress before the fate of any of these facilities is sealed. But MIT physicists have still been stunned by the recommendation. According to Robert Redwine, director of the MIT Laboratory for Nuclear Science, a recent interim report from the federal Nuclear Science Advisory Committee (NSAC) underscored the critical importance to the nation of

maintaining small university accelerators.

The Bates accelerator was specifically cited for its contribution to US nuclear physics. "The NSAC worked hard to develop a long-range research strategy that was optimal for the nation," Redwine added. "There is no indication that the House subcommittee members paid much attention to this review."

MIT would be hit particularly hard by the House subcommittee plan. The money it receives from DoE supports 22 graduate students, 6 faculty members, and 30 graduate students from other universities, as well as an operating staff of 122. The facility is used by 250 scientists from 50 institutions around the world. "Educational value alone is not enough to keep facilities going," says Redwine.

Stanley Kowalski, the director of Bates, admits that he and his colleagues need to do more to explain the attributes of their nuclear physics facilities. "Most Congressmen are not scientists, and we'd like to educate them," he said. "We want them to take a careful look at what they're proposing to see if it's really best for the country."

Steve Nadis

Tax problems threaten cuts in young scientist grants

►tive to the disbandment proposal. The department is likely to be cut back considerably more than Secretary O'Leary would like; last week, the House energy and environment subcommittee, chaired by Dana Rohrabacher, passed an authorization bill that would cut DoE's civilian research and development activities next year by a quarter, from \$5.3 billion to under \$4 billion.

DoE laboratory directors are now working fervently with O'Leary to improve their relations with the department along the lines of proposals contained in one part, Appendix B, of the report by the industrialist Robert Galvin into their future (see *Nature* 373, 463; 1995) — the second-best option, according to Galvin. Without such a course of action, observes one laboratory director, reform of the laboratories becomes "a wide open ball game for the crazies".

But the effectiveness of reform will be hindered by the gulf between O'Leary and the laboratories, where she is admired and despised in almost equal measure. While many DoE scientists complain of her poor attention to detail and lack of commitment to sacred nostrums, her political and presentational skills have improved the department's once-wretched public image.

Indeed, a poll published in the *Wall Street Journal* last week found that the public opposed abolition of DoE by 63 to 25 per cent. Nevertheless, the department's fate will be decided not by the public, but by the Senate — the fulcrum of power on many issues between an angry House and a complacent administration.

Meanwhile, the Rohrabacher bill set out a worst-case scenario for the DoE's 1996 budget, which will be finalized between now and September. The bill would slash funding for solar and renewable energy research from \$420 million to \$200 million; nuclear energy research from \$190 to \$120 million and fusion research from \$370 million to \$230 million.

It would, however, seek to protect basic science by raising support slightly for basic energy sciences and high-energy physics. "If you're a wolf caught in a trap and you need to cut a leg off to escape, best make sure it's the right leg," asserts Rohrabacher, a fiery conservative who clearly enjoyed seeing off attempts to soften his spending cuts.

The House Science Committee is passing unusually detailed authorization bills this year. This is an attempt to reassert its influence on the appropriations subcommittees, which have the last word on budgets. The attempt has the backing of the House leadership and the appropriations bills are likely to contain broadly similar numbers.

The Rohrabacher bill would also eliminate seed funding of \$6 million for US participation in Europe's Large Hadron Collider (LHC) project. But Walker says he and John Myers (Republican, Indiana), chairman of the energy appropriations subcommittee, both support the LHC and will seek to restore this money. **Colin Macilwain**

Munich. The European Commission may have to reduce by 500 the number of fellowships it can offer to young scientists intending to study in another European Union (EU) state. The commission told the council of research ministers in Luxembourg last week that it may have to extend a decision taken last December — to reimburse those holding fellowships under its new Training and Mobility of Researchers (TRM) programme who have to pay tax — to the entire duration of the four-year programme.

The problem has arisen because some member states do not exempt EU research grants from tax and social security deductions. But the compensation will come out of the commission's research funds, and its extension would, therefore, mean a 10 per cent reduction in the number of fellowships that can be offered.

The TRM programme is part of the EU fourth Framework programme, and will provide ECU260 million (US\$343 million) in fellowships to young scientists wishing to work for up to two years in another member state between 1995 and 1998.

It follows the previous Human Capital and Mobility programme, which ran into

predicted the difficulties in administering its grants fairly (see *Nature* 361, 196; 1993).

In response, the commission set up a committee last year to establish a payment system tailored to national pay scales, as well as the tax and social security conditions in individual member states. Delegates from each state have proposed figures that would allow grant recipients to receive take-home pay equivalent to that of local scientists.

These figures, which have been informally approved by member states, will be used when the first TRM grants from the programme are paid in October. In addition to the basic pay, the commission will provide an additional sum of about ECU200 a month to compensate for the inconvenience of living abroad, and around ECU10,000 a year to the host institute for bench fees.

When the new programme was put before research ministers for their approval in December, it was agreed that, in order to avoid a delay in its launch, the commission should cover the cost of taxes for the first year of the programme. The council also asked the commission to set up a committee of taxation experts to recommend a longer-term solution.

Presenting its conclusions to the council of research ministers last week, the taxation committee proposed that the system of individual compensation be continued until the end of the current programme, as it saw no short-term prospect of persuading all member states to harmonize their tax laws on this issue. But it suggested that the commission should work towards ensuring harmonization by the beginning of the fifth Framework programme in 1999.

The solution — which is likely to be approved by the council of research ministers — saves the commission from further embarrassment on the issues. But because the compensation money must be found from its own resources, the commission estimates that it will be able to offer 10 per cent fewer fellowships.

Although many scientists will benefit from the new scheme, it may take time to win the confidence of those who find EU programmes more trouble than they are worth. David Colquhoun, for example, professor of pharmacology at University College London, applauds the goal of promoting mobility of scientists within Europe. But he adds: "the bureaucracy involved [getting EU grants] is widely laughed at. It's ten times as much work to get funds from the EU as from national sources."

Despite such reservations, the previous Human Capital and Mobility programme attracted nearly 12,000 applications, 30 per cent of which were funded. **Alison Abbott**



difficulties because it offered a single rate for grants regardless of the host country's tax and social security rules. As a consequence, young scientists working in countries such as France, where they were required to pay such taxes, found themselves in financial difficulties, while those working in countries such as the United Kingdom or Italy sometimes ended up with higher pay than their departmental heads.

Further complications arose because bursaries were not accompanied by bench fees, while no allowances were made for differing overhead requirements between host institutes, or for fluctuating exchange rates. As a result, the commission came in for some harsh criticism for not having

Germany minister backs genome research

Heidelberg. Germany's research minister, Jürgen Rüttgers, speaking this week at the German Cancer Research Centre (DKFZ), promised that Germany will catch up with the rest of the world in genetics research by the end of the decade.

But Rüttgers disappointed both journalists and scientists by announcing that he was not able officially to launch Germany's new genetics research programme, expected to be worth DM50 million (US\$72 million) a year for eight years, as planned.

The delays have apparently been caused by last-minute concern within the research ministry over the level of industry's contribution to the financing of the programme.

Despite excellent research facilities and generous funding, public resistance stemming from the Nazi period has kept Germany behind countries such as the United States, the United Kingdom and France in genetics research.

But Rüttgers said that public opinion is now changing, and that although the potential use of genetic engineering in the food and agricultural industries continues to be viewed with great suspicion, there is good

acceptance of the value of such techniques in medicine.

He added that there are also signs that German industry has at last got its foot in the door. Although it has only four genetically engineered medicinal products on the market, compared to 80 in Japan and 130 in the



Rüttgers: 'public views are changing'.

United States, Germany now has 22 in the development stage.

In presenting the concept of the national genetics research programme — whose official launch is now due within the next few weeks — the research ministry has been careful

to stress that its goal is to develop diagnostic tools and therapies for human diseases, even though in practice such applications are quite distant from the programme's main aim.

That is to develop a national genetic

resource centre to provide scientists with genetic probes for localizing and identifying genes (see *Nature* 375, 175; 1995). German scientists, in academic or industrial institutes, will receive them free of charge, on condition that they provide all of their data to the centre; a charge will be made for non-German scientists and industry.

Critics have argued that Germany may be offering too little, too late, and that it has little hope of catching up other countries — certainly not within five years. But Rüttgers expressed optimism, pointing out the recent change of public attitudes, as well as a relaxation of regulations on genetics research, which had previously required a disheartening amount of paperwork from researchers.

Rüttgers refused to be drawn on whether he would shift funds from other areas into genetics research. But he emphasized that he is a staunch defender of basic research, pointing out that when he took office last November, he became embroiled in an acrimonious debate about whether Germany should transfer funds from basic to applied research, but failed to find any reason for doing so.

Alison Abbott

Physicist will sue Internet providers over 'libellous' remarks

London. A British physics lecturer who last week accepted out-of-court damages in Britain's first Internet libel case now intends to issue libel writs against two Internet providers in the United States and Canada.

Laurence Godfrey, who secured a "substantial" payment from Philip Hallam-Baker, a nuclear physicist formerly working at CERN in Geneva, over remarks about his professional work posted on Usenet, is preparing legal action against a university in Canada and a private Internet access-provider in the United States. Both organizations, he claims, have refused requests to censor remarks posted from their sites that he believes are defamatory.

"I have written to the organizations on several occasions," says Godfrey. "If they had an apology [from those who had made the original remarks] and promised not to allow the remarks to be repeated on their networks, I was happy to let the matter drop. This has not happened, which leaves me with no option to sue for libel in order to clear my name." Both organizations, which Godfrey is refusing to name, have said they will defend any proposed action.

Nick Braithwaite, a lawyer with London-based international law firm Clifford Chance, says that although the resolution of the UK case does not establish any legal precedent — in particular whether his remarks should be judged by the laws of libel or slander — one explanation of the decision by Hallam-Baker's lawyers to with-

draw is the thought they would lose.

Nonetheless, Godfrey's suit against Hallam-Baker was based on personal libel. The absence in either the United States or Canada of legislation on Internet libel makes the prospects of legal action against Internet-providers uncertain.

Editorial interference on the Internet is a sensitive issue. Most US bodies that provide access to the Internet regard any interference as a violation of free expression (see *Nature* 374, 297; 1995).

Internet libel cases, though rare, are not unknown. In 1991, a federal judge in New York decided that the online service CompuServe could not be sued for libel when it unknowingly let defamatory comments remain on one of its bulletin boards. The judge is understood to have made his ruling on the basis that CompuServe, which exercised no editorial control, was an innocent party, and could not be treated as the publisher of the material.

But in another case last month, the US investment bank Stratton Mount scored a partial victory in its \$200-million libel case against the proprietary online service Prodigy for allowing allegedly scurrilous remarks about the bank to be posted on a bulletin board by a subscriber.

In an interim, pre-trial judgement in May, the New York State Court judge ruled that Prodigy's status during the hearing would be as publisher of the material, and not as an innocent carrier of information.

According to Robert Carolina, also of Clifford Chance, "the reasons for this were that Prodigy openly advertised that they exercise editorial control over material available to subscribers". But he adds that an equally important issue — whether the statement can be considered defamatory — has yet to be resolved.

Internet libel has had less exposure in the United Kingdom. But one university said it would not wait for a court decision before removing material posted on a bulletin board that appeared defamatory. "If something looked defamatory, says Roland Rosner, director of the Information Systems Division at University College London, "I'd consult lawyers first and, if need be, remove it on the grounds that Internet access at UCL is for academic purposes only."

But private operators consider policing the Internet to be an unprofitable tier of administration. "It's just too complicated," says Steve Kennedy of Demon Internet, an organization that provides its subscribers with unrestricted access to Internet. "I can't employ hundreds of people to wade through half a gigabyte of news each day," he says.

The UK government will introduce draft legislation later this year to protect 'innocent' Internet providers from being sued for defamatory material passing through or held on bulletin boards — "assuming they did not know of its defamatory nature", according to a spokesman for the Lord Chancellor's Department.

Hesan Masood

Uproar greets new blood scandal indictment

Paris. France's long-running contaminated blood affair last week reached what may be a turning point, with the indictment on charges of "collusion of poisoning" of Jean-Baptiste Brunet, a leading AIDS epidemiologist.

Not only has the indictment itself provoked widespread protest, but it has brought to the boil simmering concern about the French legal system's handling of the affair.

The judiciary has not made public the grounds for the charges against Brunet, who now heads the European Centre for Epidemiological Monitoring of AIDS in Paris. Indeed, some claim it has not yet found specific evidence against him, and has indicted him on the basis of his overall role in the affair, intending to narrow the charges later.

Brunet's role stems from a period he spent at the Department of Health in the early 1980s, during which he wrote several memoranda drawing attention to the probable contamination of French blood supplies with the AIDS virus.

His indictment has been welcomed by the Association des Polytransfuses, a pressure group that claims to represent those who received contaminated blood. In a strongly-worded statement, the association criticizes what it describes as an "indecent media campaign by Brunet's fan club" to ridicule the legal system, and attacks Arnaud Marty-Lavauzelle, the president of patient-group AIDES, which has supported Brunet.

But Brunet's indictment has been swiftly and vigorously condemned by scientists, physicians and AIDS patient associations. In his defence they claim he showed remarkable prescience in recognizing and publicizing the threat of AIDS in the early 1980s.

An "astonished" Peter Piot, director of the United Nations' AIDS Programme, said "there is no way we can reproach Brunet for not having done his job". At that time, said Piot, "very few realized the public health issues, and the scientific and medical data were far from being as clear as today".

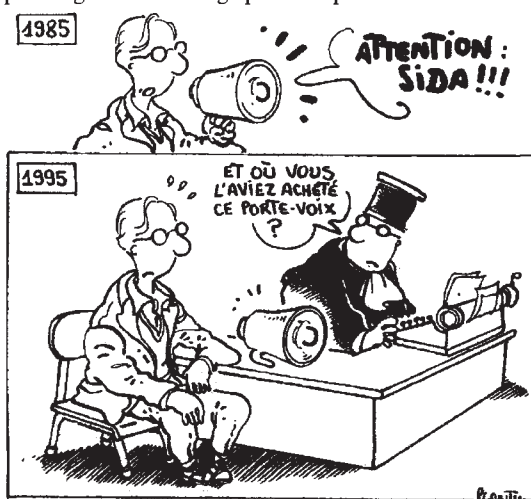
Moreover, Luc Montagnier of the Institut Pasteur, a member of the group that co-discovered the AIDS virus in 1983, also points out that Brunet, who at the time was a young physician at the Claude Bernard Hospital in Paris, held only a contractual post at the health department. He thus had little power to influence decisions, says Montagnier, especially as AIDS was then considered to a minor public health problem.

Montagnier says Brunet was one of a small group of physicians and researchers who in the early 1980s formed an AIDS study group that met every Saturday morning in Montagnier's office. They were regarded with contempt by much of the scientific and medical community,

who considered that they were overreacting.

A feeling that the judiciary has gone too far in indicting Brunet has prompted many to question the legal system's overall handling of the affair, and for the first time to propose alternative strategies. "The essential thing now", says Piot, "is to finally calm the debate."

Brunet's indictment is the latest in a series of controversial legal decisions. Last summer, for example, the Paris bar association publicly warned the French legal system against allowing public opinion to take



As *Le Monde* sees it: "And where did you buy this megaphone?" a judge asks indicted AIDS researcher.

precedence over the law in its prosecution of individuals involved in the affair (see *Nature* 370, 319; 1994).

The association's stand followed the indictment of Michel Garretta, the head of the National Centre for Blood Transfusion, and Jean-Pierre Allain, its research director, on charges of "poisoning" for the same actions for which they were given prison terms in 1992 — allegedly supplying haemophiliacs with clotting factors infected with the AIDS virus. The association argued that a second trial infringed both the law and human rights.

But Montagnier and others argue that the judges are just doing their job, and that blame for the current situation lies with the Supreme Court of Appeals, whose ambiguous interpretation of the term "poisoning" left the door open to the new charges.

The court's interpretation also prompted the current investigation — which has been widened to include an alleged delay in introducing routine screening of blood for HIV — in which similar charges have been brought against three former ministers and several of their advisers.

Last week, *Le Monde* predicted that the "only escape" from the current situation was for the entire affair to be thrown back to the plenary session of the Supreme Court of

Appeals, France's highest legal body, for review. "How can we understand the legal system's current approach?", the paper asked, asserting that judging the affair had become impossible.

Another concern is that the current spate of legal proceedings may not lead either to the fair exercise of justice, or to a balanced explanation of malfunctions in the French health system in the early 1980s.

"We mustn't let the 'tree' of Garretta hide the forest of all the physicians involved in transfusion," says François Gremi, a professor of public health at the University of Montpellier and a member of the High Committee for Public Health — equivalent to the US Surgeon General's Board. "It's a much more general problem; the whole of the profession failed to do its job."

A report to the government by the same committee, on policy issues raised by the use of human products in medicine, reaches similar conclusions. "The sense of our report is not to judge whether Garretta and Allain are innocent, but to point out that if we focus on individuals, we risk forgetting that the errors depended on an entire system," says Gremi.

Gremi admits that Garretta and his colleagues may have played an important role. But he adds that to avoid the same thing happening again "we must understand the profound causes".

Wider issues have been neglected, says Gremi, partly because attention has been focused on the clotting factors case. "The haemophiliac issue was a drama but not a scandal," he says, adding that the non-application by doctors of a 1983 circular — ironically, instigated by Brunet — warning of the need to exclude blood donors from high-risk populations was a "much bigger problem".

For the first time, Montagnier has also publicly criticized the legal proceedings, saying that while they may have helped to clarify the situation, they have become "ridiculous". Montagnier says he refrained from speaking out before on the handling of the affair because of the "irrational" mood of public opinion stirred by media coverage of the plight of haemophiliacs infected by contaminated blood products. He says that, rather than continuing indefinitely with the trials, the need "is to start a dialogue among scientists, physicians, haemophiliac associations and journalists to find a way forward".

In a statement released shortly after his indictment, Brunet hoped the move would at least "provide an opportunity to clarify completely the entire affair, including the contamination of the haemophiliacs, the introduction of tests for screening of blood, and the collection of dangerous blood, in particular from prisons". **Declan Butler**

Physicist-turned-ecologist to take top UK advisory post

London. Last Monday — the day after England beat Australia in the rugby world cup — John Major, the British prime minister, redressed the balance by announcing the appointment of Robert May, an Australian-born physicist turned mathematical ecologist, as his chief scientific adviser.

May, who is currently Royal Society research professor at the University of Oxford and Imperial College, London, will succeed Sir William Stewart in a post which also includes responsibilities as head of the Office of Science and Technology.

A combative and frequently outspoken researcher, May is widely known for a keen intellect which can occasionally seem impatient with those promoting ideas that cannot be backed by persuasive quantitative arguments.

But he is also known as a dedicated researcher with unusually broad academic interests, and as an individual who has taken an active interest in issues ranging from career structures for young scientists, to the scientific basis of preserving biodiversity.

"I think it is an excellent appointment. Bob is an individual who can think quickly on his feet and knows his own mind, and is also firmly in touch with the issues facing the scientific community," says one of his academic colleagues.

May, who is 59, was born in Sydney, where he gained a PhD in theoretical physics. He taught at Sydney University from 1962 to 1973, when he was appointed professor of biology — and later Class of 1877 Professor of Zoology — at the University of Princeton.

It was over this period that he turned from physics to theoretical ecology. May says he "blundered into the golden age of mathematical biology", with a particular interest in topics such as the population biology of infectious diseases.

He now suggests that his scientific background may have provided him with skills that will stand him in good stead in his new post. "In my work, I tend to take a very complicated real-world situation, form a view of what is the most important thing going on, put it in precise terms, and then say: what have I learnt about the system as a whole?"

May gained experience in research administration as Princeton's vice-president for research, a position which carried responsibility for research policy and administration. As such, he says, he experienced "in some sense the same relation to the president of the university as the chief scientist does to the British prime minister."

Given that May has had relatively little direct involvement in conventional UK science policy circles — news of his possible

appointment generated some surprise as a result (see *Nature* 374, 583; 1995) — his track record at Princeton may have weighed in his favour. May points out with some pride that during his term of office, he not only doubled the volume of research at the university, but increased "from 2 to 12 per cent" its income from industrial sponsors."

Since being appointed to his current post in 1988, May has become deeply involved in many topics at the intersection of biology and public policy. These range from debates over biodiversity — he was for a time a trustee of the World Wide Fund for Nature, and is currently chairman of the trustees of the Natural History Museum — to the epidemiology of AIDS.

His own intellectual interests have been closely concerned with the development of chaos theory and nonlinear dynamics, including their application to fields ranging from theoretical immunology to the behaviour of stock markets. May says that he hopes to be able to spend a small amount of his time to continuing such work. "It will give an added dimension and strength to my new job if I can remain actively engaged in experiments."

But he also has few illusions about the size of the task that he is taking on, with a portfolio that includes topics ranging from the continuation of the government's technology foresight exercise to the use of animals in experiments. May remains discreetly silent about his views on such topics. But he is full of praise for his predecessor. "I have the characteristic of saying what I feel about things, and in this case, that is not too difficult," he says. "Bill Stewart has done a wonderful job."

Others suggest some of his likely priorities. "He is, for example, very supportive of young scientists, and keenly aware of the need for universities to put in place career structures for their researchers," says Roy Anderson, a fellow professor of zoology (and epidemiologist) at Oxford.

Another colleague claims that "there is a certain lack of numeracy in government, and I hope that Bob is able to redress the balance." Anderson describes May as "someone who always wants to confront theory with observation." His new post will provide him with ample opportunity to do so.

David Dickson



May: supportive of young scientists.

ESA science head warns of threat to future missions

Paris. Roger Bonnet, director of the space science programme of the European Space Agency (ESA), warned this week that future space missions may have to be delayed or cancelled if the United Kingdom insists on a 25 per cent reduction in the costs of the programme when Europe's space ministers meet in Toulouse in October.

Speaking at a press conference during the Paris air show, Bonnet said that if ministers approved cuts of such magnitude, there would be no "miracle solution", and ESA would be forced to delay by five years one of the 'cornerstone' missions — large missions budgeted to cost around ECU625 million (US\$825 million) — within the current Horizon 2000 programme.

Two possible candidates in such a scenario would be either the X-ray telescope XMM, which is scheduled for 1999, or the Rosetta mission, due to be launched around 2003 with the aim of landing instruments on the surface of a comet around 2013.

Given the amount of work that has already gone into the first of these, Rosetta could be the more vulnerable. Bonnet adds that a five-year delay could make a mission scientifically uncompetitive, and it would probably be cancelled instead.

ESA had previously also hoped to obtain a four to five per cent increase in the budget of Horizon 2000+, the science programme that will run for ten years from 2006, in order to pay for what would be the space agency's first venture into the field of fundamental physics.

This cornerstone mission would be aimed at detecting gravitational waves by launching six satellites millions of kilometres apart and keeping them connected by laser beams. Bonnet says that if the project is judged to be feasible, ESA will now look for "entrance fee" funding instead from the particle physics community.

Declan Butler

Indian appointments

New Delhi. India last week named **Valangiman Ramamoorthy**, a prominent physicist who is at present director of the Institute of Physics in Bhubaneswar, Orissa State, as head of the Department of Science and Technology. He replaces **P. R. Rao**, who retires later this month.

The government has also appointed a new director general for the Council of Scientific and Industrial Research, the country's main scientific agency with responsibility for more than 40 laboratories. **Ragunath Mashelkar**, a polymer scientist who trained in the United Kingdom, will take over from **Shri Krishna Joshi**, who is retiring on 1 July.

Space station survives test of Congress' political will

Washington. The international space station has begun its annual voyage through the US budget process with a show of support from House Republicans, whose solid backing seems likely to give the project its easiest ride through Congress for several years.

Last week, the House space and aeronautics subcommittee, chaired by James Sensenbrenner (Republican, Wisconsin), overwhelmingly approved a multi-year, \$13-billion authorization bill which, although unlikely to pass into law, will send a clear signal of Congress's intention to see the controversial project through to the end.

The subcommittee rejected, by 18 votes to 3, an amendment from Tim Roemer (Democrat, Indiana) to kill the space station. Almost all the subcommittee Republicans are freshmen and enthusiastic budget-cutters. But under the watchful eyes of Sensenbrenner and Bob Walker (Republican, Pennsylvania), chair of the Science Committee, all of them backed the bill and opposed the Roemer amendment.

Ralph Hall (Democrat, Texas), ranking minority member of the subcommittee and a strong supporter of the station, warned Sensenbrenner of the risk of bringing such a bill to the House floor "in the current climate". But given the solid support of Newt Gingrich (Republican, Georgia) and other Republican House leaders, and the strong party discipline of their troops, most observers see little risk in such a vote this summer. The station retains the strong support of the Clinton administration and of most senators.

George Brown (Democrat, California), ranking minority member of the Science Committee, says he will continue to push for some protection of the total budget of the National Aeronautics and Space Administration (NASA) in exchange for his support of the multi-year authorization bill. But his supporters concede he has little leverage to pursue his goal.

The authorization bill proposes that US funding for the space station should remain constant at \$2.1 billion a year within an overall NASA budget which, under the House budget resolution (see *Nature* 375, 168; 1995), will fall from \$14.5 billion this year to \$11.7 billion by the year 2000.

During the subcommittee debate, many justifications were put forward for the space station. Hall asserted it was "the only hope" for "people wasting away with cancer in hospitals". Dave Weldon (Republican, Florida) sought biblical inspiration for the station, quoting the biblical proverb that "where there is no vision, the people perish", while Steve Stockman (Republican, Texas), made the more straightforward argument that his wife was employed on the project.

The reality is that Boeing's contract to build the US portion of the station will result in much of the work being strategically placed in hard-pressed aerospace facilities in Texas and California. Neither political party is going to take that work away in the run-up to a presidential election, and the international partners concerned about the US's commitment can sleep soundly at nights.

Colin Macilwain

Scripps agrees to revised plan for ocean experiment

San Diego. The California Coastal Commission is meeting in Carmel today (15 June) to decide whether to approve the Californian part of a controversial project designed to study whether sound waves can be used to measure global warming in the ocean.

The chances of approval have been increased by an agreement last week between the Scripps Institution of Oceanography in San Diego and a number of environmental groups to turn the \$35-million project into a two-year research programme that will initially concentrate on the effects of sound on marine mammals and only incidentally collect climate data.

Full-scale climate studies would begin only after an environmental review. The shift in emphasis follows concern expressed by environmentalists that the experiment, originally planned purely as way of using the speed of sound through water to measure ocean water temperature, might damage marine animals.

The experiment is known as Acoustic Thermometry of Ocean Climate (ATOC), and will be carried out by an international team of scientists headed by Walter Munk at Scripps. It involves using two transducers in about 1,000 metres of water — one off the coast of northern California, and another off the Hawaiian island of Kauai — to send out coded signals lasting about 20 minutes.

ATOC was supposed to start last year with one transducer in Monterey Bay, a national marine sanctuary. But in March, following criticism from environmentalists, Scripps reduced the planned number of emissions, and moved the transducer to the slopes of a dormant undersea volcano 89 kilometres off the coast.

The new schedule of experiments has now been approved by organizations such as the Environmental Defense Fund, the Natural Resources Defense Council (NRDC) and the Sierra Club Legal Defense Fund. Ann Notthoff of NRDC says the new project is entirely different from that first proposed. "This is now *bona fide* marine mammal research," she says. "Marine mammalogists now control the sound source for two years."

But other organizations still object. "The only thing that has changed is that they call it an extended marine mammal research programme when it isn't," says Ray Chuan, of the Kauai Friends of the Environment.

Although they will be primarily gathering data on the effects of the sounds on marine mammals, researchers will also begin collecting data on water temperatures. "It is not optimized for our work," said Andrew Forbes, the project director. "But we will certainly get some good data." **Joel Shurkin**

Two satellites prove better than one

London. The first combined images from the European Space Agency's remote sensing satellites ERS-1 and ERS-2 (left), should dispel any lingering doubts about the decision to operate the two in tandem. The mission exploits the unique ability of each satellite's 10-centimetre resolution synthetic aperture radar to compare two pictures of the same area taken at different times. The result is a digital three-dimensional map of central Italy.

Consecutive ERS-1 and ERS-2 images, although recorded at the same hour on different days, reveal a difference in soil moisture that reflects changed weather conditions,



indicated by the greenish colour of the land. The colours of the sea correspond to different wind and current conditions during the three acquisition periods. Black indicates calm areas at all acquisition dates. E. M.

UK fraud verdict prompts moves on ethics...

London. Attempts by the British scientific research establishment to prevent scientific fraud were given new urgency last week when the General Medical Council struck off the medical register a consultant, Malcolm Pearce, who had been found guilty of publishing two fraudulent papers in the *British Journal of Obstetrics and Gynaecology*.

The Medical Research Council (MRC) is drafting a comprehensive document on good research practice, to be released next year, that will cover the topics of scientific fraud and 'whistleblowing'.

Martin Kemp, head of research policy development at MRC, says that "there is scope for laboratories to do more internal screening". The MRC is planning a clear sanctions package, coupled with rigid procedures for inquiring into scientific fraud designed to safeguard whistleblowers.

Meanwhile, the Royal College of Physicians (RCP) is expected to announce this week plans to set up an independent body to regulate medical research. According to its registrar, David London, the college would like to see the creation of a body able to commission inquiries and receive complaints, similar to those that already operate in Scandinavia. "Where there is crime, one needs a degree of policing," says London.

Pearce, a consultant obstetrician at St George's Hospital in London, was found guilty of 12 charges of professional misconduct — including two of scientific fraud — over two papers published in August 1994.

In the first paper, co-authored with his head of department, Geoffrey Chamberlain, who resigned last week as president of the Royal College of Obstetrics and Gynaecology, Pearce claimed that he had carried out the world's first successful transplantation of an ectopic pregnancy.

The second paper contained an entirely fabricated report on 191 women suffering recurrent miscarriage. When questioned, Pearce later claimed none of the records of those included in this study could be found.

Issuing its verdict, the GMC, which regulates medical practice in the United Kingdom, condemned Pearce's "deliberate inclusion of false or misleading data in medical research" as dishonest and intolerable. The council described as "incalculable" the consequences for public confidence in the integrity of research.

In an effort to restore public confidence, the GMC emphasized the need to devise and operate effective safeguards, especially in the area of 'gift authorship' — the practice whereby senior academics add their names to research papers with which they have had little direct involvement. The council suggests that "all individuals named in a research paper must have made an intellectual contribution and be able to verify the raw data".

Stephen Lock, a former editor of the *British Medical Journal* who is an expert on the study of scientific fraud, describes the Pearce case as "a tragedy waiting to happen". He was impressed, however, by the swift and efficient manner in which the case — which lasted only eight months from beginning to end — was handled.

Lock, who advises the RCP on dealing with fraud cases, admits that medical

research will inevitably produce a small percentage of "bad apples". But he adds that "peer review cannot — and can't be expected to — detect these", especially as "the whistleblower usually suffers more than the miscreant". He also supports calls for gift authorship to cease, comparing the practice to "saying that you helped Shakespeare write Hamlet because you lent him a pencil".

Sara Abdulla

... as US reviewer resigns over slur

San Diego. A New York epidemiologist has resigned from a grant-reviewing committee of the US National Institutes of Health (NIH) over the treatment of a grant application from a research team at the University of Alabama at Birmingham, which was last month found guilty of stealing a key idea used in the application from a former research student.

Karla H. Damus of the Albert Einstein College of Medicine has resigned from the maternal and child health research committee of the National Institute of Child Health and Human Development, on which she had served for more than three years. Her resignation stems from her participation in the 1992 review of an application to renew a \$4.8-million grant from the Alabama team, which has been studying congenital cytomegalovirus (CMV) infections in newborn children.

Last month, the team was found guilty of using research results stolen from another scientist improperly to obtain federal grants (see *Nature* 375, 270; 1995). Details about the review process emerged during the federal lawsuit, in which Pamela A. Berge, an epidemiologist, had sued Robert F. Pass — head of paediatric infectious diseases at the university, and the principal investigator for a five-year grant seeking a vaccine against CMV — and three Alabama colleagues for stealing results obtained while she was completing a thesis and for using them to win grants improperly from the institute.

A federal judge in Baltimore, Maryland, issued a \$1.9-million judgement against the university and the four researchers. Damus — who acknowledges having met Berge previously — claims that a spurious allegation that she had a conflict of interest was used by the Alabama team in 1992 to get its application re-reviewed after the review committee had not given it a sufficiently high rating to receive funding.

In March 1992, two days after the committee's decision, Pass wrote to institute officials claiming that Berge had influenced the review either through Damus or another review committee member. The institute subsequently

convened a second committee of different reviewers, but that committee also gave the application an unfundable rating.

A third review by yet another committee finally gave the Alabama application a sufficiently high rating for it to be funded in May 1993. Institute officials say that the committee was unaware of the earlier decisions. But Damus claims that the successive reviews raise questions about the institute's policies for ensuring grants are evaluated on the basis of quality alone, and also for protecting reviewers who might be inappropriately criticized by grant applicants.

Damus claims that in 1992 she was unaware of Berge's dispute with the Alabama team, and says she did not learn about the conflict allegation — or the other two reviews of the grant — until one of Berge's attorneys contacted her last autumn. "I am furious over this," says Damus. "It is unconscionable that no-one at the institute would pick up the phone and call about such a serious allegation. I cannot be a reviewer until they set policies and guidelines to protect reviewers."

Charlotte Catz, the institute's project officer on the Alabama grant, acknowledged that three successive reviews were "unusual" and had never previously occurred in her 18-year tenure. She describes Damus as "an excellent reviewer", adding that it is regrettable that she was upset.

Although the institute never investigated Pass's conflict of interest allegations, Catz says that the additional reviews were "the most efficient way" of handling the matter. But Damus describes the institute's actions as a way of "covering up" a "really effective old boys' network" for reviewing and funding grants.

Pass has been unavailable for comment since the lawsuit, believed to be the first case in which a jury has ruled that scientists used research results stolen from another scientist to obtain federal grants. Despite last month's jury verdict, attorneys for Alabama and the researchers deny any impropriety, and are seeking to have the verdict set aside.

Rex Dalton

Paradox of placebo effect

SIR — When testing for the effectiveness of a drug in ameliorating a disease, it is common practice to compare the treated group to a control population, matched in all essential respects, to whom is administered a placebo containing substances that are presumed to be inert. The placebo-controlled trial is an integral part of evidence-based medicine. Although the nonspecific effects of placebos are widely studied, the possibility that the chemicals used as placebos may have specific effects has received virtually no attention.

The US Food and Drug Administration sets no regulations on the constituents of placebos, and any guidelines are at best informal. Astonishingly, no systematic efforts are made to ensure the inertness of placebos: there is nothing validating the placebo standard against which other agents are measured. Further, the drug companies funding the trials control the placebo ingredients.

The identity of the placebo and fillers used with the experimental drug are rarely stated in scientific studies. In one exception to this practice, several early papers exploring the use of cholesterol-lowering agents to curb heart disease did in fact name the placebos used: olive oil in one case¹, and corn oil in another². Mono- and poly-unsaturates such as olive oil and corn oil are now widely known to decrease low-density lipoproteins³, so that with hindsight these agents may not have been inert with respect to the outcome studied. Indeed, it was noted in one such study² that the rate of cardiac mortality was lower in the placebo group than expected.

How can we be sure that placebos are free of specific effects? Few if any agents are truly inert, and placebos are given systematically over prolonged periods. Even substances that are not absorbed, such as methylcellulose (which reduces cholesterol), can have significant effects.

Placebos used across trials may differ: such differences are among many factors that may underlie outcome differences in otherwise similar trials. Placebos used across trials may, on the other hand, be similar, or have similar effects: thus we cannot rely on meta-analyses to cancel differing specific effects of placebos. In addition, it is felt by many that the correct solution to the thorny problem of irreconcilable smaller studies is a single very large trial rather than a meta-analysis. Indeed, in some fields the evidential basis of a population-based treatment regimen rests on the outcome of a single significant trial. The Scandinavian Simvastatin Survival Study⁴ is the lone cholesterol-lowering trial, among many, that showed improved overall mortality with cholesterol reduction during the time of treatment, even in the secondary prevention population; it is

cited as proof of the benefits of cholesterol reduction therapy. Clearly, small effects by a placebo may be pivotal when clinical practice is determined by the results of a single very large trial, whose large size is necessitated by small effect size of the treatment.

Until recently the possibility of small effects of placebos accruing over long periods may not have been of serious concern because most trials sought large effects in small subject populations studied over relatively short times. Large-scale trials (or aggregate analyses of trials) involving thousands of individuals studied over many years and seeking small effects (see, for example, ref. 5) are a modern phenomenon. It is in this setting that small beneficial or harmful effects of placebos could be significant. An apparent positive, negative or null effect of a drug may instead be the consequence of a negative, positive or same-direction effect of the placebo.

What should be done? First, it is essential that all studies should state the composition of the placebos as well as all fillers included with the drugs. Studies should be carried out to test for possible specific effects of placebo agents, even though this process will be fraught with difficulty. Meanwhile, in interpreting the results of prior trials, we should be aware that possible specific effects of placebos represent a potential confounding factor that may be vital to interpreting the study results.

The foundation of evidence-based medicine is undermined by the absence of evidence that placebos are inert. It is paradoxical that there is no standard of evidence to support the standard of evidence.

Beatrice A. Golomb

Department of Medicine,
University of California, Los Angeles,
Los Angeles, California 90024, USA

1. Acheson J. & Hutchinson E. C. *Atherosclerosis* **15**, 177–183 (1972).
2. Research Committee of the Scottish Society of Physicians, *Br Med. J.* **4**, 755–784 (1971).
3. Grundy S. M. & Denke M. A. *J Lipid Res* **31**, 1149–1172 (1990).
4. Scandinavian Simvastatin Survival Study Group, *Lancet* **344**, 1383–1389 (1994).
5. Lau J. et al. *New Engl. J. Med.* **327**, 248–254 (1992).

Vatican confusion

SIR — Recent correspondence (*Nature* **372**, 124; 1994 & **373**, 278; 1995) demonstrates the confusion that the BC/AD baseline of the calendar can generate. Now it appears that even the Pope is confused. In his Apostolic Letter *Tertio Millennio Adveniente* (Libreria Editrice Vaticana, 1994), he proclaims (p. 16) that *Anno*

bismillesimo Magnum idcirco erit Iubilaeum ("The Great Jubilee will be therefore on the 2000th year"), and refers (p. 25) to the *Magnum Iubilaeum exeunte altero millennio* ("The Great Jubilee at the end of the second millennium") and (p. 40) to the *Magnum Iubilaeum quod annum bismillesimum concludet* ("The Great Jubilee which closes the 2000th year"); but then he states (p. 41) that the *Christiani invitantur ut se tertii millennii Magni Iubilaei ab initium incohandum expediant* ("The Christians are invited to prepare themselves for the Great Jubilee beginning at the start of the third millennium") and concludes (p. 50) by asking the Virgin Mary to be like a guiding star *Christianis procedentibus tertii millennii in Magnum Iubilaeum* ("for the Christians proceeding toward the Great Jubilee of the third millennium").

Until p. 40 the Pope correctly assigns the year 2000 to the second millennium; but after p. 40 he assigns it incorrectly to the third millennium. Following a procedure initiated by Pope Alexander VI in AD 1500, the Great Jubilee proclaimed by John Paul II will start on Christmas Eve of the year 1999 and terminate on Christmas Eve of the year 2000, seven days before the beginning of the third millennium.

Ian Elliott

Dunsink Observatory,
Dublin 15, Ireland
Cesare Emiliani
International Academy of Sciences,
276 Kelsey Park Circle,
Palm Beach Gardens,
Florida 33410, USA

Side by side

SIR — The juxtaposition of two leading articles in a recent issue of *Nature* (**374**, 392; 1995) was mildly entertaining, and mightily scaring.

In the left-hand column ("Equity and addiction") it was suggested that perhaps, in order to "save some people's lives", "users of [marijuana, tobacco and alcohol] should require a medical licence to do so".

In the right-hand column ("Murder and the metro") it was stated that "[t]he outlook for strictly technical control of the ingredients of the gas [sarin] are [sic] not bright, given the damage inflicted upon liberal societies by protective regulations that are themselves intrusive and thus illiberal". (My italics.)

Should one infer that, in your view, regulations are less "illiberal and intrusive" when they strictly regulate the private lives of individuals than when they strictly regulate corporations?

Marta Monteleoni

28 Siddons Court,
39 Tavistock Street,
London WC2E 7NT, UK

Towards traps for cold molecules

A scheme for trapping molecules in strong infrared laser beams has been described theoretically by two chemists. Experimental proof awaits. When it comes, molecular spectroscopy will be transformed.

COOLING atoms to low temperatures is now almost child's play. The simplest thing to do is to start with ions, decelerate them in an electrostatic field and then catch them in an electromagnetic trap which may be either a geometrical arrangement of electrostatic and magnetic fields or simply a pattern of electromagnetic forces as generated by mutually perpendicular Helmholtz coils on a very small scale. Then it is a matter of robbing the ions of kinetic energy, for which purpose laser light detuned from a resonance line seems to be the most effective means. Temperatures of a millikelvin are accessible. The benefits to spectroscopy are immense.

Molecules, sadly, are much less easily trapped and cooled. Ionization is often a prelude to dissociation, the electrons in an ionized molecule may otherwise be redistributed into different states and there is the complexity of rotation and vibration as well as excitation to contend with. Even so, there are some experiments that suggest that molecules may be aligned with each other by high-intensity laser pulses provided that they are sufficiently brief to avoid dissociation — a picosecond or so.

That seems to be the starting point for a remarkable article by two Harvard chemists, Bretislav Friedrich and Dudley Herschbach, which essentially outlines a scheme for trapping and cooling molecules of most simple kinds (*Phys. Rev. Lett.* **74**, 4623–4626; 1995). The remarkable feature of this study is that it is entirely a theoretical study.

In molecular cooling, the alignment of molecules is the first need. If a molecule has a dipole moment, simple electrostatic fields should do the trick. Of course, the accuracy of the alignment will turn on how much kinetic energy an ensemble of molecules has available for oscillation about the dipole axis. One technique for doing this is to cool the molecules (enclosed in a good vacuum) by the vapour from a pool of liquid ^3He , which can rob an ensemble of dipolar molecules of their rotational kinetic energy.

But what if the molecules carry no permanent dipole moment? That is the point from which Friedrich and Herschbach take off. Whether or not molecules have permanent dipoles, they are polarizable in electric fields. What better, to induce that condition, than an intense laser beam? But how do molecules behave in such conditions? On the face of things, it is simply a matter of solving a Schrödinger equation

for the angular rotation of a molecule dealt with as if it were a rigid rotator — vibrations are, in other words, neglected.

The problem is tractable if the frequency of the radiation is greater than the reciprocal of the duration of the laser pulse. Then the effects of whatever permanent dipole moment there may be are averaged out. The results of the calculation are interesting only if the polarizability of the molecule is asymmetric, which is usually the case. For a diatomic molecule, for example, the polarizability along the internuclear axis and in the perpendicular direction will almost certainly be different. But if the polarizability in two directions should be equal, the rotation of the molecule would correspond to the rotation of a molecule in the absence of an electric field, which provides a neat way of classifying the states of the molecule in the presence of a field in terms of the angular momentum of the molecule and its projection along the axis.

The eventual outcome is a calculation of the statistical properties of a collection of molecules in the intense electric field of a powerful laser beam in the infrared region of the spectrum. Evidently the degree of alignment of molecules at any temperature will increase with the rotational constant (or moment of inertia) of the rotating molecule. The greater the temperature, the less good the alignment will be. But for anisotropic molecules and at lower temperatures, the mean-square of $\cos\theta$, where θ is the angle between the orientation of the molecule and its axis of polarizability, is arbitrarily close to unity; alignment should be straightforward.

In due course, these calculations will find their way into the textbooks of quantum mechanics. The states of a molecule in an electric field strong enough for the polarization to be significant are called "pendular" states. They should provide spectroscopists with a field day; the selection rules determining which states can be converted into others by emission or absorption of infrared radiation are shown to be different from those familiar to molecular spectroscopists.

How likely is it that these conditions can be achieved in a real laboratory? Friedrich and Herschbach applaud the neatness of the experiments at Saclay in which D. Normand, L. A. Lompre and C. Cornaggia first demonstrated alignment in CO molecules (*J. Phys. B* **25**, L497; 1992), in which liquid ^3He was first

used for cooling. The difficulty will be that of making laser beams intense enough not merely to win good alignment of molecules but to trap them in very small places.

That should work because polarizable molecules should gravitate to regions in which the intensity of the field is greatest. So much is true in classical physics, but the quantum mechanical calculations bear out the expectation. It should then suffice as a trap for molecules to arrange that the laser should concentrate its energy into a volume of 10^{-9} cm^3 or $10\text{ }\mu\text{m}$ in size. They reckon that it would not be necessary to use a pulsed laser, and that a continuous-wave laser delivering a few kilowatts of power would suffice if its energy were concentrated in the trapping volume by sufficiently accurate mirrors.

By the same test, the use of liquid ^3He for robbing the gas molecules of translational and vibrational energy should suffice. The technique is to keep the liquid helium in a cryostat and to allow its natural vapour pressure to do the cooling. Obviously it is necessary that the vapour pressure should be high enough to carry away the excess energy. According to Friedrich and Herschbach, the energy states of the trapped molecules should be lower than the energy of escape by the equivalent of 1 K.

They are even able to make predictions of the trapping process for three molecules — CO, CS_2 and C_{60} . They reckon that they should be able to achieve concentrations between 10^8 and 10^{13} cm^{-3} with suitable lasers and coolant gas. These are not very different from the concentrations the atom trappers boast about.

There is even the possibility of further cooling of an ensemble of trapped molecules by allowing those with the highest energy to escape periodically. 'Evaporative' cooling on this model is used in cooling trapped atoms, and should work just as well for atoms.

So why have Friedrich and Herschbach not carried out the measurements whose outcome they predict with such exquisite certainty? There are some obvious reasons why they should hang back; theoreticians are not always natural experimentalists. But then it is an act of great generosity on their part to have given some group somewhere a detailed protocol that could almost certainly be a successful grant application overnight.

John Maddox

Applying a genetic cantilever

Marco E. Bianchi and David M. J. Lilley

INTERACTIONS between DNA and proteins lie at the very heart of development. In the latest issue of *Cell*¹ Marius Clore and his colleagues zoom in near the atomic level to the complex cascade of switches that control the expression of key genes during eukaryotic development. Using nuclear magnetic resonance, they have solved the solution structure of a complex between a DNA octamer and the DNA-binding domain of SRY, the protein factor that is responsible for determining male gender in mammals².

SRY is a member of the HMG class of DNA-binding proteins that are characterized by their effect on DNA structure³. A striking feature of the SRY-DNA complex is that the DNA is bent and twisted, in marked contrast to DNA in complex with most repressor proteins, in which the protein recognizes a target binding sequence but leaves the DNA largely unperturbed. This effect of SRY illustrates the inherent deformability of DNA, and substantiates proposals that HMG-box proteins function as architectural components to ply and mould DNA⁴. The local structural deformation of the double helix induced by SRY might mediate effects at a distance through the mechanical displacement of DNA segments (and associated factors) at either side of the point of flexure.

The HMG domain was first identified as duplicated 80-amino-acid regions in the abundant non-histone nuclear protein HMG1. Sequences with significant similarity were then found in a variety of transcription factors, including SRY, LEF-1 and UBF and in more than 100 other DNA-binding proteins (reviewed in ref. 5). These proteins appear to be subdivided into classes depending on their sequence selectivity, from the essentially nonspecific HMG1 and 2, through UBF which is partially specific, to SRY, LEF-1 and TCF-1, for which proper consensus DNA-binding sites can be defined. What all these proteins have in common is their marked effect on DNA structure, and an affinity for distorted DNA structures such as four-way junctions⁶ and *cis*-platinum adducts⁷. HMG proteins bend their target DNA, retarding electrophoretic mobility and promoting cyclization rates⁸, and can even participate as auxiliary factors in site-specific recombination reactions^{8,9}. This has led to the idea that the HMG box is an all-purpose DNA bender/wrapper/looper which can be recruited for a variety of DNA transactions, including transcription, repair and recombination — a kind of eukaryotic version of the IHF and HU proteins of *Escherichia coli*.

Three earlier NMR structures of HMG

boxes of the non-sequence-selective class^{10–12} exhibited the same general fold with minor differences, in which three α -helical segments formed an L-shaped structure stabilized by a hydrophobic core. The longer part of the 'L' comprised the helical C-terminal region and an extended N-terminal section. In the SRY-DNA complex¹ the overall fold is closely similar, but the angle of the 'L' is more obtuse. The concave surface of the protein is perfectly located in the minor groove of the DNA where it buries a large surface area. To achieve this, the DNA is distorted by a succession of positive roll angles and bent back by 70 to 80 degrees. The protein is locked down on to the DNA by a series of interactions with the backbones of both helices. The minor groove is extensively widened to accommodate the protein, accomplished principally by a wedge of five residues that prise apart the central region of the DNA.

A factor in the bending of the DNA is the intercalation of an isoleucine residue at the centre of the structure^{1,13}. The extent of bending can vary between different HMG proteins: for example, LEF-1 induces a considerably greater bending than SRY⁹. The protein structure itself might influence the degree of bending, with target DNA being bent more by HMG boxes having a more acute-angled 'L', or the protein may adjust this angle on binding to the DNA; in this regard it would be interesting to know the structure of free SRY in solution. What is clear is that very different binding surfaces in an expanded minor groove can bend and unwind DNA in a similar manner, because the overall shape of the SRY-DNA complex is strongly reminiscent of that formed by the TATA-box-binding protein with its target DNA sequence^{14,15}.

It is not apparent why four-way junctions should be a universal target for HMG-box proteins. No convex surfaces are presented in the folded junction, but the real target may be the unfolded junction found at low salt concentrations, or some other conformation induced by protein binding. Another outstanding issue is the origin of the sequence specificity of SRY and the LEF-1/TCF-1 families and of the sequence indifference of the HMG1 family. A domain-swap experiment showed that a chimaeric protein consisting of the long arm of TCF-1 fused to the short arm of HMG1 retains the binding and bending ability of TCF-1 (ref. 16). Clore and colleagues¹ have identified a number of contacts to DNA bases distributed along the entire binding surface that would be expected to mediate specificity. Many of these positions are altered in

the non-sequence-selective HMG-box proteins.

The structure of the SRY-DNA complex offers a rare opportunity for molecular interpretation of mutations that cause clinical effects. Natural mutants of human SRY are associated with abnormal sexual development and all involve the HMG box. A few mutations can be inherited, which means that sometimes they can be recovered from normal males, and so must be less disruptive than mutations that cause malformation of the gonads and cannot be transmitted. Interestingly, all inherited point mutations map to sites that do not interact with DNA; DNA binding overall is not grossly altered^{17,18}, suggesting by default that the stability of the protein could be affected. Some *de novo* mutations map to sites that would severely disrupt protein packing and four involve residues that contact DNA and would be expected to reduce the DNA affinity of SRY.

One mutation, involving an isoleucine substitution for a methionine at residue 64, is remarkable in that it does not appreciably reduce the affinity or the specificity for the DNA target but affects the extent of DNA deformation, as detected by the alteration of the electrophoretic mobility of the complex¹⁸. The methionine involved is part of the wedge that kinks DNA by unstacking the central base pairs, so this conservative substitution might reduce the surface of contact between protein and DNA. Thus, a modest alteration of the SRY-induced DNA structure can translate into a severe disruption of developmental patterns and a serious clinical condition. □

Marco E. Bianchi is in the Department of Genetics and Microbiology, University of Milano, via Celoria 26, 20133 Milano, Italy. David M. J. Lilley is in the CRC Nucleic Acids Structure Research Group, Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK.

1. Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. *Cell* **81**, 705–714 (1995).
2. Sinclair, A. et al. *Nature* **346**, 240–244 (1990).
3. Lilley, D. M. J. *Nature* **357**, 282–283 (1992).
4. Grosschedl, R., Giese, K. & Pagel, J. *Trends Genet.* **13**, 94–100 (1994).
5. Bianchi, M. E. in *DNA-Protein: Structural Interactions* (ed. Lilley, D. M. J.) (IRL, Oxford, in the press).
6. Bianchi, M. E., Beltrame, M. & Paoonessa, G. *Science* **243**, 1056–1059 (1989).
7. Pili, P. M. & Lippard, S. J. *Science* **256**, 234–237 (1992).
8. Pauli, T. T., Haykinson, M. J. & Johnson, R. C. *Genes Dev.* **7**, 1521–1534 (1993).
9. Giese, K. et al. *Cell* **69**, 185–195 (1992).
10. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C. & Norman, D. G. *Nucleic Acids Res.* **21**, 3427–3436 (1993).
11. Weir, H. M. et al. *EMBO J.* **12**, 1311–1319 (1993).
12. Jones, D. N. M. et al. *Structure* **2**, 609–627 (1994).
13. King, C.-Y. & Weiss, M. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11990–11994 (1993).
14. Kim, Y. C., Geiger, J. H., Hahn, S. & Sigler, P. B. *Nature* **365**, 512–520 (1993).
15. Kim, J. L., Nikolov, D. B. & Burley, S. K. *Nature* **365**, 520–527 (1993).
16. Read, C. M. et al. *EMBO J.* **13**, 5639–5646 (1994).
17. Harley, V. R. et al. *Science* **225**, 453–456 (1992).
18. Pontiggia, A. et al. *EMBO J.* **13**, 6115–6124 (1994).

Rules for the manipulation of polyketides

John Mann

GENETIC engineering meets organic chemistry in the paper by McDaniel, Ebert-Khosla, Hopwood and Khosla on page 549 of this issue¹. The authors describe the latest step in their work with the bacterium *Streptomyces coelicolor*, the aim of which is to formulate rules for the rational design of natural products derived through polyketide pathways.

Streptomyces coelicolor is a member of the order Actinomycetales, which contains some of the most primitive species that inhabit this planet. Yet these humble soil microorganisms produce a large number of biologically important and clinically useful products. Well-known examples include antibacterial compounds such as the tetracyclines and erythromycin, the antitumour anthracyclines, and the immunosuppressant FK506, and the antiparasitic avermectins. These secondary

metabolites have diverse chemical structures that fall into two broad categories: polycyclic aromatic compounds such as daunomycin (one of the anthracyclines; **1** in Fig. 1) and macrocyclic compounds such as erythromycin A (**2**). All of these are, however, derived from polyketide intermediates which are assembled by a family of enzymes called polyketide synthases (PKSs) (reviewed in refs 2,3).

It is only 40 years since Birch proposed (and then demonstrated) that polyketides are constructed from acetyl- and malonyl thioesters; and it is less than 20 years since Staunton, Simpson and others, using the stable isotopes deuterium and ¹³C in conjunction with nuclear magnetic resonance, began to reveal the subtleties of the biosynthesis of the aromatic polyketide metabolites (for reviews see refs 4,5). This research demonstrated that the structural diversity was a result of the different numbers and types of acyl units involved, and pointed to the various possibilities for enzyme-mediated reduction, dehydration, cyclization and aromatization that existed. Manipulation of the sequence or specificity of these processes to produce novel secondary metabolites was an obvious aim of subsequent research.

The key to success was an understanding of the genetics of the various bacterial PKSs. Elegant studies by the groups of Katz, Hopwood and Hutchinson⁶⁻⁸, amongst others, established that there are two main types of the enzyme. The modu-

lar type has discrete multifunctional enzymes controlling the sequential addition of thioester units and their subsequent modification to produce macrocyclic compounds such as erythromycin A; unique active sites are used for each chemical step. In contrast, the aromatic PKSs are composed of a small number of separate proteins, the active sites of which are used iteratively for the assembly and

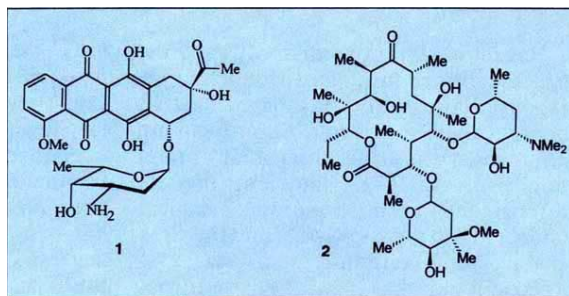


FIG. 1 Products derived from polyketide intermediates. (1) The polycyclic aromatic compound daunomycin. (2) The macrocycle erythromycin A. The first has antitumour activity, the second is an antibacterial agent.

functional-group manipulation of the polyketide chain.

These aromatic PKSs contain a set of three genes which encode a so-called 'minimal PKS'. This includes a ketosynthase which catalyses the condensation between the acyl thioester species, an acyl transferase, a chain-length factor and an acyl carrier protein which acts as an anchor for the polyketide chain during the various chemical manipulations. Such a minimal PKS can produce the simple metabolite known as SEK4 (**3** in Fig. 2) from eight C-2 units, but to produce the reduced compound SEK34 (**4**) the additional activities of a ketoreductase and an aromatase are required. With a minimal PKS in the presence of a ketoreductase, an aromatase and a cyclase enzyme, additional ring formation takes place between C-4 and C-15 (numbering from the thioester end of the polyketide), and the anthraquinone shown in (**5**) is the end product. The challenge was then to obtain a suitable combination of a minimal PKS, a ketoreductase, an aromatase and a cyclase that would assemble and transform progenitors with various chain lengths

into a range of novel metabolites.

Earlier studies by McDaniel *et al.*^{9,10} had shown that a strain of *S. coelicolor*, from which the *act* aromatic PKS gene cluster (and other genes) had been excised, could act as a convenient host system. They have now gone further, and introduced a number of plasmids containing combinations of PKS, ketoreductase, aromatase and cyclase genes from various organisms into the modified *S. coelicolor* to produce a range of transformed strains with new aromatic PKS components. These were then screened for the production of both known and unknown metabolites. From the variations in the number and types of rings produced, and in the functional groups they possessed, the authors have been able to propose a set of tentative rules for polyketide design.

Many of the rules are not unexpected, such as the requirement for a ketoreductase if ketoreduction is to occur, or that the overall chain length is dictated by the minimal PKS. But the identification of the combination of gene products required for the various cyclizations is both fascinating and potentially very useful. For example, whereas the first ring in unreduced polyketides aromatizes without the need for an aromatase, a functional aromatase is required once ketoreduction has occurred (see Fig. 2). In addition, with a minimal PKS, ketoreduction at C-9 will induce cyclization between C-7 and C-12, but ketoreduction at C-7 induces cyclization between C-5 and C-10.

The rules must be regarded as provisional, but they nonetheless open the way for the rational design and production of completely novel metabolites. As McDaniel *et al.* note¹, the extent of molecular diversity will be controlled by the number of degrees of freedom for each nascent polyketide chain. These will include chain length, location of the cyclizations and the level of aromatization. Thus far the group has only manipulated PKSs that produce chain lengths up to 24 carbons; but naturally occurring polyketide metabolites of up to 28 carbons have been isolated, so these PKSs should be available for manipulation. Add to this the possibility of using building blocks other than acetate and malonate (propionate and methylmalonate for example), or a combination of PKS genes with those controlling the oxygenases and glycosyl transferases involved in the final stages of biosynthesis, and the scope for the production of structures becomes enormous.

Whereas the work reported by Mc-

- McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *Nature* **375**, 549–554 (1995).
- O'Hagan, D. *The Polyketide Metabolites* (Ellis Horwood, Chichester, 1991).
- Simpson, T. J. *Chem. Ind.* 5 June, 407–411 (1995).
- Simpson, T. J. *Chem. Soc. Rev.* **4**, 497–522 (1975).
- Garson, M. J. & Staunton, J. *Chem. Soc. Rev.* **8**, 539–561 (1979).
- Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. *Science* **252**, 675–679 (1991).

- Bibb, M. J. *et al.* *EMBO J.* **8**, 2727–2736 (1989).
- Fernandez-Moreno, M. A., Martinez, E., Boto, L., Hopwood, D. A. & Malpartida, F. *J. Biol. Chem.* **267**, 19278–19290 (1992).
- McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *Science* **262**, 1546–1550 (1993).
- McDaniel, R., Ebert-Khosla, D., Hopwood, D. A. & Khosla, C. *J. Am. Chem. Soc.* **116**, 10855–10859 (1994).
- Cortes, J. *et al.* *Science* **268**, 1487–1489 (1995).

Daniel *et al.* wholly concerns aromatic PKSs, exciting results have also been obtained with the modular PKSs. In last week's *Science*¹¹, Leadlay, Staunton and their co-workers described their manipulation of the genes of *Saccharopolyspora erythraea*, a species that produces erythromycin A. They have successfully repositioned a chain-terminating cyclase

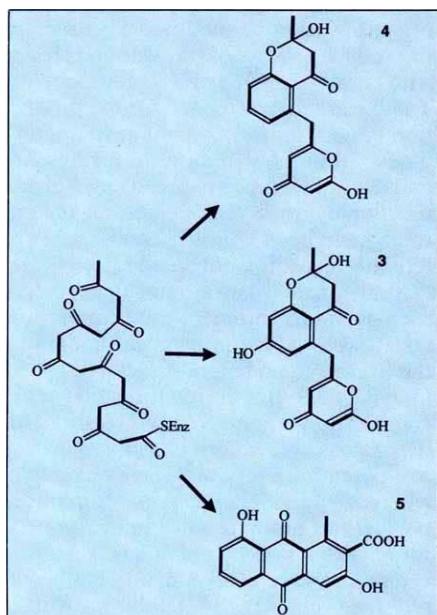


FIG. 2 The major products of an octaketide (eight C-2 units are emphasized) in the presence of (3) an act minimal PKS alone; (4) an act minimal PKS, act ketoreductase and act aromatase; and (5) an act minimal PKS, an act ketoreductase, an act aromatase and an act cyclase.

domain to the carboxy terminus of the first multienzyme. The normal enzyme catalyses the synthesis of a partially reduced triketide, which is then passed on to the next multienzyme for further addition and processing of methylmalonate units. The mutant strain made one major product — a triketide lactone (and no trace of erythromycins) — clearly the result of premature chain termination and cyclization. The scope for production of novel cyclic metabolites is obvious, and this research nicely complements the work of McDaniel *et al.*

Nature has spent millions of years 'experimenting' with PKSs and in the process has provided us with a rich variety of polyketide metabolites. With the growing understanding of the biosynthetic pathways which lead to these structures, it should be possible within a matter of years — rather than millennia — to produce a plethora of 'unnatural' products. It would be amazing if some of these did not possess clinically useful activity. □

John Mann is in the Department of Chemistry, University of Reading, Whiteknights, Reading RG6 6AD, UK.

Lines drawn in epitope wars

John M. Coffin

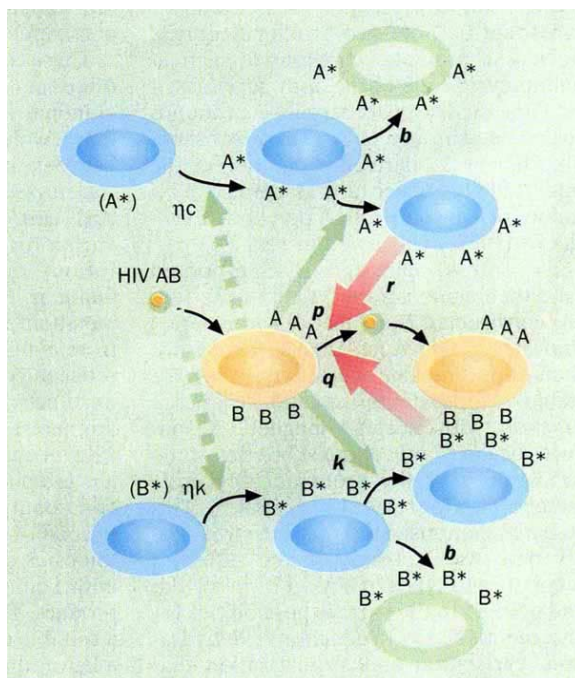
THE role of the host immune response in controlling infection by human immunodeficiency virus and the reaction of the virus to it are complex and controversial topics. It is generally accepted that variants relatively resistant to neutralization by antibodies appear during infection, but the importance of these to the actual course of infection is less obvious. Even more uncertain is how cytotoxic T cells (CTLs) generated in the cellular immune response contribute to controlling infection and forcing antigenic diversity.

Recent observations that virus and infected cells are turning over much more rapidly than previously thought¹⁻³ have brought these issues into sharp focus by removing the misconception of HIV infection as a relatively indolent process through the extended period of clinical latency. With the knowledge that the virus is replicating constantly at all stages of infection, and that it is transmitted from one infected cell to another every one to two days, we are now forced to confront directly the issue of how the infection is kept under such firm control for so long. The paper by Nowak *et al.*⁴, appearing on page 606 of this issue, while it does not directly answer these important questions, provides a frame of reference for seriously exploring them.

The study consists of two parts: observations on the CTL response to previously defined HIV epitopes in infected individuals, and the development of a mathematical model to describe these responses. Two patients (007 and 020) are presented, both of them haemophiliacs infected for some time with HIV. Both began the observation period with fewer than normal CD4⁺ T cells, and these declined further during the almost 5-year study. In patient 007, who remained well, only a single reactive epitope could be detected in the virus-infected cells. Although two sequence variants of this sequence were present, as well as CTLs reactive against each, the situation was quite

stable, with little fluctuation over the time of observation. The other patient, who died of AIDS, behaved quite differently, displaying reactivity against three epitopes of the virus in a temporally fluctuating pattern. Also fluctuating was the proportion of HIV sequence variants at each of the epitopes. The pattern observed suggested the possibility of oscillation of the immune response to specific sites linked to virus 'escape' mutants at the various sites.

A mathematical model of the CTL response was constructed to explore the forces that might underlie this variation in patterns. The model starts from the concept shown in the figure. Each cell infected with a virus that contains a pattern of epitope variants (A and B in the figure) stimulates a T-cell response to each of these variants (green arrows) and is in turn killed by these cells (red arrows). This leads to a set of simple differential equations describing the rate of T-cell activation and stimulation as a function of the number of infected cells carrying each epitope and the rate of infected-cell killing as a function of the number of activated T cells specific for each variant epitope. The



The concept used by Nowak *et al.*⁴ to model the immune response against HIV. In the example shown, infected cells (yellow nuclei) present two epitopes (A and B) to CTLs (blue) and induce or activate their response (A* and B*) with rate constants c and k , respectively. The activated cells kill the epitope-carrying cells with rate constants of p and q . Unstimulated cells undergo first-order decay with a rate constant b . The virus reproduces with a rate constant r . CTL are produced by activation from a pool of precursor cells at rates ηc and ηk (see ref. 4).

rate of replication of each variant virus is thus its inherent rate of replication (black arrows) reduced by the rate at which cells infected with it are killed by CTLs reactive against all epitopes it carries. Not all such variants will have the same inherent replication rate; nor will they stimulate or be killed by T cells at the same rate. Thus, for each variant virus, a separate equation with distinct parameters (each shown in bold in the figure) is needed to describe its net replication. Similarly, each variant epitope needs a separate equation to describe the CTL response to it. An additional parameter is also required to describe the decay of the CTL response when unstimulated.

Although the differential equations are fairly simple in concept, complexity mounts rapidly. The simple situation discussed by Nowak *et al.*⁴ of two epitopes, each with two antigenically distinct variants, requires solution of eight differential equations and assignment of values to no less than fourteen parameters. It is perhaps not too surprising that a wide variety of curves can be generated depending on the values selected. These curves can vary from seemingly stable equilibria of antigen and CTL to complex oscillating patterns, with the response shifting — seemingly without warning — from one epitope to another.

Different classes of curves with different general properties at various sets of values can be described, however. For example, there is a prediction of competition between different epitopes, leading in some cases to immunodominance where a response against only one epitope can be detected despite the presence of others. This will happen when the response against one epitope is somewhat greater than against others. In this case, an equilibrium between the cells displaying this epitope and specific CTLs will be reached at numbers of infected cells too low to stimulate a response from the other epitopes. This equilibrium can be maintained only if the virus remains homogenous at the dominant epitope; that is, if escape mutants of adequate replication ability and reduced immunogenicity do not arise. In the latter case, the ever-shifting pattern of response and counter-response rules.

This variability of predicted responses of CTL and virus brings us back to the two patients described, who appear to represent one of each type. The virus in patient 007 appears bonded to a single epitope, suggestive of the immunodominant model in which competitively successful escape mutants cannot arise, whereas patient 020 shows the complex pattern predicted for shifting immunodominance in the face of emerging and disappearing variant viruses. Obviously two patients are far too few to form a basis for any convincing

conclusion, but the authors suggest that the more leisurely course of disease progression in patient 007 may reflect the more successful response against a relatively unchangeable epitope.

A number of questions naturally arise. First among these is the ultimate cause of AIDS. The system described by Nowak and colleagues is oscillatory, but does not progress. The authors accommodate this problem by incorporating a previous model⁵ and propose that the immune response requires help from CD4⁺ T cells, and that, as the virus kills CD4⁺ T cells, the immune response must inevitably lose the race. Second, the model presumes that the immune response — particularly the CTL arm — is the principal factor limiting HIV replication *in vivo*. An alternative (which I myself favour) is that HIV infection *in vivo* is limited by the availability of target cells, and that the actual role played by the immune response during the clinically latent phase may be rather minor. Experiments could be designed to distinguish between these important alternatives.

Finally, it must be pointed out that the value of modelling all but the simplest

aspects of HIV infection in this way is not universally recognized in the field. Critics of this sort of approach are quick to point out that the choice of which aspects of the biological system to model and the values to assign to the parameters are somewhat arbitrary. Coupled with this is an inherent imprecision in many of the biological measurements that the models are intended to explain; for example, the analysis is limited to a few epitopes conserved with model viruses. As a result, say the critics, these models are essentially untestable and of poor predictive value. But even imperfect models and imprecise data can serve to raise the awareness of biologists and mathematicians to both the problems and the potential of projecting HIV infection through equation. □

John M. Coffin is in the Department of Molecular Biology and Microbiology, Tufts University, 136 Harrison Avenue, Boston, Massachusetts 02111, USA.

1. Ho, D. D. *et al.* *Nature* **373**, 123–126 (1995).
2. Wei, X. *et al.* *Nature* **373**, 117–122 (1995).
3. Coffin, J. M. *Science* **267**, 483–488 (1995).
4. Nowak, M. A. *et al.* *Nature* **375**, 606–611 (1995).
5. Nowak, M. A. *et al.* *Science* **254**, 963–969 (1991).

PLANETARY SCIENCE

Sublime Solar System ices

William B. McKinnon

WITHIN some future human settlement on a distant icy moon, students of geology and geophysics will eventually come to the textbook chapter on some most unusual natural materials — the silicates. Luckily, the variety of surrounding ice-forming materials will help them understand the behaviour of those mineral systems so important to Earth-bound scientists of the distant past. In our time, we are accumulating the knowledge of natural ices that will fill these future books. One important step was taken when scientists from over 15 countries gathered recently* to discuss all aspects of ices in the Solar System (and some beyond).

But what exactly is an ice, as opposed to say, a rock? The colloquial meaning is water ice, which is also the master ice of the Universe, as oxygen is the most abundant of the ice-forming elements. We also expect 'ices' to melt and vaporize at relatively low temperatures. Solid-state physicists have defined ice as any molecular solid, that is, one formed from molecules that are weakly (van der Waals) bonded. For astrophysicists, this covers many species such as methane, N₂, CO, and even CO₂, SO₂ and formaldehyde. But as Bernard Schmitt (LGGE CNRS, Grenoble) pointed out at the meeting, this

excludes H₂O and its hydrogen-bonded brethren ammonia and methanol. So we are left with an imprecise definition, part common usage, part volatility criterion and part solid-state chemistry. Whether one views solid sulphur as an ice or a rock then depends on whether one has an Io-based or an Earth-based perspective.

Whatever the perspective, there are many ices and many settings. We have come a long way from J. S. Lewis's original description¹ of ice chemistry in the outer Solar System based on the simplest reduced forms of oxygen, carbon and nitrogen, H₂O, CH₄ and NH₃ (see figure). It has become increasingly evident that less reduced forms such as CO and N₂ are also important on several icy satellites and in comets. These are the original chemical forms inherited from the molecular cloud from which the Solar System was born, and in the outer solar nebula the temperatures were low enough that conversion to thermodynamically preferred reduced forms was kinetically inhibited over the lifetime of the nebula (currently estimated to be several million years)². Even fully oxidized CO₂ can be found within comets, and is a major surface ice on Neptune's satellite Triton, in addition to its well-known role in the martian polar caps. The list does not stop there. Materials in cometary bodies prob-

*Solar System Ices, Toulouse, France, 27–30 March 1995.

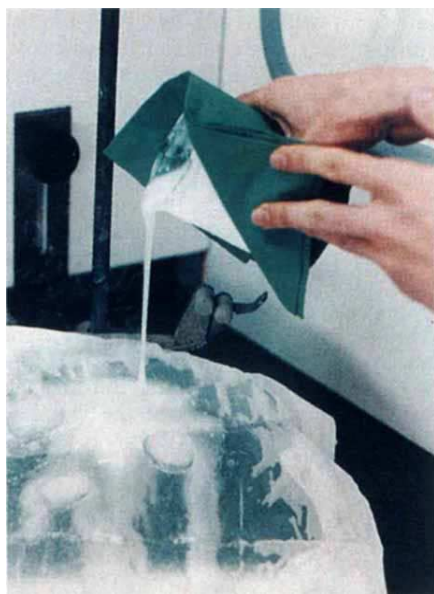
ably were accreted into icy protoplanets, of which Triton, Pluto, Chiron and various Kuiper belt (or disk) objects³ may be representative. There may also be a substantial cometary heritage in other regular icy satellites. Important cometary ices such as methanol and formaldehyde may thus be geologically and geophysically important in satellite evolution (J. Kargel, US Geological Survey, Flagstaff). Water ice itself has been identified from the Pluto–Charon system to (it is thought) Mercury, where it is presumed responsible for the strong radar reflections from the bottoms of permanently shadowed polar impact craters (D. Paige, Univ. California, Los Angeles).

New condensed phases have been found on the galilean satellites of Jupiter: in particular, molecular oxygen has been detected in high-resolution visible wavelength spectra taken from the ground (W. Calvin, US Geological Survey, Flagstaff; ref. 4). The discovery is perplexing in that at the surface temperature of Ganymede liquid or solid O₂ should evaporate immediately. In addition, the absorption features were seen exclusively on the trailing hemisphere, the side of Ganymede that is continuously exposed to bombardment by magnetospheric charged particles as the satellite synchronously turns. The O₂ may be a product of the interaction of this energetic plasma with Ganymede's surface of water ice (R. Johnson, Univ. Virginia). Similar bombardment may produce the ambient, albeit extremely tenuous, O₂ atmosphere seen on Europa⁵, another of the jovian moons.

Condensed O₂ is not seen on Europa's trailing hemisphere, and Calvin and co-workers suggested that it may be scavenged by implanted magnetospheric sulphur atoms to form SO₂. Ultraviolet spectra from the Hubble Space Telescope indeed show an absorption feature that is a good match to laboratory spectra of SO₂ (K. Noll, Space Telescope Inst.). This hypothesis for SO₂ formation has not, however, been corroborated by laboratory experiments so far, which leads to an intriguing alternative — endogenous SO₂ (ref. 6). As SO₂ is an important volcanic volatile on both the Earth and Io, it is plausibly present as a contaminant in the surface ice or dissolved in any subsurface fluids, even if only a relic of an earlier volcanic or hydrothermal stage in Europa's evolution. On the other hand, the band strength varies with longitude and is the deepest on Europa's trailing hemisphere, which is clearly a magnetospheric effect. Possibly, preferential sputtering of water ice on the trailing hemisphere leads to the gradual enrichment of SO₂ there⁶.

There is, interestingly, new evidence that Europa remains internally active. A rift-like feature in the satellite's ice shell has been found near its south pole

(R. Pappalardo, Arizona State Univ.). The feature is geologically young and joins a similar though broader area of rifting near the anti-jovian point⁷. Both of these regions thus become part of the geological evidence that supports the hypothesis of an ice shell floating on a liquid water ocean at Europa, an ocean that does not freeze because there is tidal heating provided at the base of the ice shell⁸. The south pole of Europa will be



A possible lava in the outer Solar System: ammonia–water liquid near the 176 K peritectic. Rheologically similar to molten basalt, this material has been implicated in volcanic terrains, but not spectroscopically identified, on several satellites of Saturn, Uranus and Neptune.

imaged by the Galileo spacecraft on its way in to Jupiter this December. When the images are sent back, we may have confirmation of the rifting, further evidence for liquid water at depth and spectral maps that indicate where SO₂ or presently undetected ice species may be localized. Europa may turn out to be one of the highlights of this mission.

Another mission, still in the planning stage, may benefit from the exotic behaviour of another ice. New calculations (J. Stansberry, NASA Ames) show that as nitrogen ice cools across the β to α (hexagonal to cubic) transition temperature of 35.6 K (α -N₂ being the low-temperature phase), its infrared emissivity drops significantly. As surface temperatures in radiative equilibrium are inversely related to emissivity, this will cause the global surface temperature, buffered by this 'supervolatile' if it is abundant (as it is on Triton and Pluto), to hang up at 35.6 K during epochs of climatic cooling. Pluto's present surface temperature is close to 40 K (ref. 9), and as it retreats from the Sun it was feared that its N₂-dominated atmos-

phere would freeze out and not be available for study when and if a mission to Pluto could be mounted. This would have been a great scientific loss, but it now appears that the planet's atmosphere will be maintained.

And what of water ice itself? A detailed evaluation of mass spectrometer measurements made during the Giotto encounter with comet Halley (P. Eberhardt, Univ. Bern) has determined the deuterium/hydrogen ratio for the (sublimed) water vapour in the comet's coma to be a rather precise 316 ± 34 parts per million (p.p.m.). This is consistent with the rather broad range previously determined, 60–480 p.p.m. (ref. 10), but is also now clearly in excess of the values estimated for the solar nebula (about 25 p.p.m.), all of the giant planets and Titan (whose D/H estimates range from protosolar values to about 100 p.p.m.), and the Earth's oceans (156 p.p.m.). Isotopic exchange between water ice and deuterated hydrogen in the solar nebula has long been considered responsible for D enrichment in Solar System water, but although detailed calculations support this as a potential explanation for the terrestrial D/H ratio, the mechanism apparently cannot explain the ratio found in comet Halley¹⁰. According to current thinking, only ion–molecule reactions at low temperatures can yield such a large enrichment. These occur in interstellar environments, and large D enrichments have been observed in molecular clouds¹⁰. So Halley's ice may be unprocessed interstellar material, or at least ice that never underwent isotopic exchange with solar nebula hydrogen.

It used to be that the D/H of terrestrial water argued for a largely cometary source for the Earth's volatiles. Now, with Halley twice as enriched as sea water, and presuming that Halley is representative of comets, it appears that about 50 per cent of the Earth's water may have come from the local solar nebula, or that the original cometary inventory had a chance to equilibrate isotopically with solar hydrogen on or within the early Earth. The sources and sinks for Solar System ices are far from being settled. □

William B. McKinnon is in the Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University, Saint Louis, Missouri 63130, USA.

- Lewis, J. S. *Icarus* **16**, 241–252 (1971).
- Boss, A. P. *Nature* **375**, 13–14 (1995).
- Jewitt, D. C. & Luu, J. X. *Astr. J.* **109**, 1867–1876 (1995).
- Spencer, J. R. *et al.* *J. geophys. Res.* (in the press).
- Hall, D. T. *et al.* *Nature* **373**, 677–679 (1995).
- Noll, K. S., Weaver, H. A. & Gonnella, A. M. *J. geophys. Res.* (in the press).
- Schenk, P. M. & McKinnon, W. B. *Icarus* **79**, 75–100 (1989).
- Ojakangas, G. W. & Stevenson, D. J. *Icarus* **81**, 242–270 (1989).
- Tryka, K. A. *et al.* *Icarus* **112**, 513–527 (1994).
- Eberhardt, P., Reber, M., Krankowsky, D. & Hodges, R. R. *Astr. Astrophys.* (in the press).

Oldest terrestrial landscape

Paul F. Hoffman

THERE are two reasons why the paper by Buick *et al.* (page 574 of this issue¹) will attract attention. The more obvious one is that it tells us something about when land started to emerge from beneath the oceans. The second is that the authors' discovery of an ancient emergent block of continental crust may, with further investigation, give us a better idea of what Earth's atmosphere was like 3,500 million years ago.

On Earth, the relationship between surface elevation and area is strongly bimodal. This is a feature that is unique among the planets of the Solar System, and it stems from the existence of oceans and plate tectonics. One mode, the abyssal plains, is at 4.5 km below sea level; that is the depth of oceanic crust at thermal equilibrium. The other mode is the surface of the continents, which is kept close to sea level by erosion and sedimentation. In keeping with Archimedes principle, continental surfaces are elevated because continental crust is thicker and lighter than oceanic crust. Continental crust originates where lithospheric plates converge: the crust is forcibly thickened, denser material is occasionally removed by entrainment in the descending mantle flow, and lighter material is sometimes added by partial melting of descending oceanic crust.

A long-held view is that continental crust accumulated slowly over billions of years, in keeping with its mean age (which is less than half the age of the Earth as a whole) and with the paucity of continental crust more than 3,500 million years old. But this view assumes that continental crust, unlike oceanic crust, has not been significantly recycled. A contrary view is that continents formed early on and their present mass and age distribution merely reflect a dynamic balance between creation and destruction. The importance of the report from Buick *et al.*¹ is not that it concerns the oldest continental-type crust (it falls short by nearly 500 million years) but that it describes the oldest continental exposure surface yet discovered.

For geologists, ancient exposure surfaces are the stuff of legend. To survive the ravages of geological time, such surfaces need protective blankets of younger sediments or lavas. The interface is called an 'unconformity', an allusion to the discordance between depositional layers above and below the exposure surface. Modern geology emerged in the eighteenth century from a failed attempt to interpret rocks and landforms as products of the Biblical Flood. Encouraged by the faithful in expectation that nature would conform to scripture, the 'neptunist' inter-

pretation was a truly scientific hypothesis in that it made unambiguously falsifiable predictions: geology should record the continuous emergence of the land from a recent oceanic inundation of the same age everywhere.

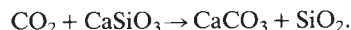
The discovery by James Hutton (1726–1797) of unconformities, where ancient sedimentary strata were forcibly upturned, eroded and then blanketed by younger deposits, pointed the way to the

"The paper by Buick *et al.* provides strong hints that a palaeosol lies beneath the 3,500-million-year-old unconformity. That is good news."

inevitable conclusion that the history of the Earth is exceedingly long and complex. Hutton's unconformity in Scotland became the geologists' shrine. It turned out to be 400 million years old, five orders of magnitude greater than the age of the Earth according to Biblical genealogy. The newly discovered¹ unconformity, which is in northwestern Australia, is almost nine times older than Hutton's.

But why report the discovery in *Nature* as opposed to the *Guinness Book of Records*? Here we come to the second reason why the paper will attract attention. There may be an ancient soil profile (palaeosol) beneath the unconformity, and palaeosol composition can be used to find out about atmospheric composition.

Where rocks emerge from the oceans, they interact with the atmosphere, forming soils. Silicate material is dissolved and carried in solution by rivers to the oceans. Silicate dissolution in soils is driven by carbonic acid rain and biologically produced carbonic and organic acids. The net effect is to remove CO₂ from the atmosphere and deposit it in the form of limestone, according to the Högbohm–Urey reaction:



CO₂ is ultimately returned to the atmosphere, in the form of volcanic exhalations, as a result of metamorphism of limestone (the reverse of the reaction above) in zones of plate convergence.

Walker *et al.*² proposed that the temperature dependence of soil 'weathering' reactions provides a negative feedback mechanism that serves to stabilize global temperature over geological time. High levels of CO₂ cause temperatures to rise through the greenhouse effect. Silicates weather faster at higher tempera-

ture, consuming larger quantities of CO₂ and causing CO₂ levels and global temperatures to drop. At low temperatures, silicate weathering and CO₂ consumption is minimal, causing CO₂ to accumulate in the atmosphere. In this way, the feedback mechanism resists large changes in global temperature (but not short-term changes such as those associated with anthropogenic modification of the atmosphere). Recent work³ confirms the sensitivity of weathering rates to changes in temperature and CO₂, and hence the strength of the feedback.

The feedback mechanism between silicate weathering and CO₂ has been invoked as a solution to the 'faint young Sun paradox', which is as follows. Geological and palaeontological evidence show that liquid water has existed on the Earth's surface for at least 3,800 and 3,500 million years, respectively. Furthermore, there are no well-established glacial deposits older than 2,450 million years. On the other hand, standard theories⁴ of solar evolution imply that the Sun's luminosity has risen smoothly from some 70 per cent of its current value 4,500 million years ago, meaning that the Earth's mean surface temperature should have been below the freezing point of water before about 2,000 million years ago. But the paradox exists only if it is assumed that the atmosphere has remained unchanged. Even if CO₂ and water vapour were the only important greenhouse gases and changes in cloudiness are neglected, an atmospheric CO₂ concentration of more than 300 times present atmospheric level could have offset the reduced solar luminosity 4,500 million years ago⁵.

Is there geological evidence for such a large secular decline in atmospheric CO₂ concentration? Grotzinger and Kasting^{6,7} have interpreted the evolution of Archaean and Proterozoic marine carbonates and evaporites in terms of declining concentrations of atmospheric CO₂. They observe that the incidence of abiotic carbonate crustose cements increases progressively with age in marine sequences older than about 1,500 million years, implying that the early oceans were in a state of increased calcium carbonate saturation. This cannot be due solely to the absence of carbonate-secreting skeletal organisms, because such cements are relatively uncommon after 1,500 million years and the first abundant skeletal organisms did not evolve until 535 million years ago.

Grotzinger and Kasting^{6,7} also note that, before about 1,500 million years ago, calcium sulphate is generally absent from the sequence of minerals precipitated during progressive evaporation of sea water. Before about 2,200 million years ago, the paucity of sulphate evaporites may be related to the anoxic state of the oceans and atmosphere. Grotzinger and

Kasting^{6,7} suggest an additional factor: that the abundant precipitation of calcium carbonate cements before about 1,500 million years ago may have reduced the calcium-to-bicarbonate ratio of sea water. They infer that calcium was exhausted during progressive evaporation before the calcium sulphate saturation field was reached. So far so good for the proponents of CO₂ as the buffer against long-term global temperature change.

This brings us back to the discovery of Buick and colleagues¹, for another means of constraining the composition of the ancient atmosphere is through the study of palaeosols developed beneath unconformities. Palaeosol data have been used, for example, to infer that free oxygen began to accumulate in the atmosphere 2,200 ± 300 million years ago⁸. Regarding atmospheric CO₂, however, Kuo *et al.*⁹ have argued that the absence of siderite (FeCO₃) in palaeosols from 2,200 to 2,750 million years ago limits concentrations to below about ten times the present atmospheric level. If, as Kuo *et al.*⁹ conclude, atmospheric CO₂ could not have been as high as required by the model of Walker *et al.*², the feedback mechanism between silicate weathering and CO₂ cannot be the primary means by which the Earth's climate is stabilized against increasing solar luminosity.

The paper by Buick *et al.* provides strong hints that a palaeosol lies beneath the 3,500-million-year-old unconformity. That is good news. The palaeosol is developed beneath the sequence containing the oldest known microfossils¹⁰. It apparently developed on both felsic and mafic igneous rocks, and subsequent metamorphism is mild (lower-greenschist grade). If problems that commonly beset the interpretation of palaeosols can be overcome, such as alteration by geologically recent weathering or groundwater flow, the oldest known exposure surface may provide a window on the composition of the early atmosphere and the possible role of silicate weathering in keeping the Earth "an abode suitable for life". □

Paul F. Hoffman is in the Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

Another DDT connection

Richard M. Sharpe

As we celebrate the half-century of the end of the Second World War in Europe, it will come as a surprise to many readers that the pesticide DDT played a part in the Allied victory. At the time, DDT was hailed as the new war weapon of the Allies¹ because of its ability to kill the lice responsible for the spread of typhoid. Only 20 or so years later was it realized that DDT was environmentally disastrous, as portrayed in Rachel Carson's *Silent Spring*. Now, 50 years on, comes

sion a detailed report⁴, the findings of which are to be discussed in the Danish parliament in the autumn.

If these trends are true (and the increase in incidence of testicular cancer is beyond dispute), the cause (or causes) must lie in changes in our environment or lifestyle over the past 50 years. The most strongly argued case has been that human exposure to a range of weakly oestrogenic chemicals, which are ubiquitous in the environment, is to blame. This case is

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Spraying it about — in 1946, whole districts of Athens were sprayed with DDT from low-flying aircraft in attempts to halt the spread of a cholera epidemic by flies.

another twist to the story, with the discovery by Kelce and colleagues (page 581 of this issue)² that *p,p'*-DDE, the main metabolite of DDT in the body, is a potent anti-androgen. The fact that this discovery comes at a time of increasing concern about adverse changes in male reproductive health^{3,4} is a remarkable coincidence. Or is it more than this?

There is evidence of a substantial fall in average sperm counts in normal men⁵ and a progressive increase in incidence of testicular cancer⁶ over the past half-century. There is other, although more equivocal, evidence that abnormal development of the penis (hypospadias) and mal-descent of the testes (cryptorchidism) have also increased in incidence⁴. It has been hypothesized that all of these changes are interrelated and that they may have a common origin in fetal and/or neonatal life³. But the veracity of the data on falling sperm counts continues to be disputed⁷, although a recent study of 1,300 fertile semen donors from the Paris region showed that sperm counts have dropped on average by around 2 per cent each year over the past 20 years⁸. In Denmark, where the lifetime risk of a male developing testicular cancer is now nearly 1 per cent, the government has been sufficiently concerned to commis-

sion a detailed report⁴, the findings of which are to be discussed in the Danish parliament in the autumn. If these trends are true (and the increase in incidence of testicular cancer is beyond dispute), the cause (or causes) must lie in changes in our environment or lifestyle over the past 50 years. The most strongly argued case has been that human exposure to a range of weakly oestrogenic chemicals, which are ubiquitous in the environment, is to blame. This case is based, first, on evidence that exposure of man and animals to inappropriate levels of oestrogens during fetal and neonatal life will result in a spectrum of male reproductive disorders similar to those listed above^{3,4}. Second, all these oestrogenic chemicals have been introduced into the environment in large amounts during the past 40–60 years⁴. They include certain detergent breakdown products and various plastics additives and stabilizers. However, it has long been known that one isomer of DDT (*o,p'*-DDT) and related chlorinated pesticides can also exert oestrogenic effects (so too can some modern

1. Buick, R. *et al.* *Nature* **375**, 574–577 (1995).
2. Walker, J. C. G., Hays, P. B. & Kasting, J. F. *J. geophys. Res.* **86**, 9776–9782 (1981).
3. Brady, P. V. & Carroll, S. A. *Geochim. cosmochim. Acta* **58**, 1853–1856 (1994).
4. Gilliland, R. L. *Global planet. Change* **1**, 35–55 (1989).
5. Kasting, J. F. *Science* **259**, 920–926 (1993).
6. Grotzinger, J. P. in *Controls on Comparative Platform and Basin Development* (eds Crevello, P. D. *et al.*) 79–106 (Society of Economic Paleontologists and Mineralogists, Tulsa, 1989).
7. Grotzinger, J. P. & Kasting, J. F. *J. Geol.* **101**, 235–243 (1993).
8. Holland, H. D. & Beukes, N. J. *Am. J. Sci.* **290**, 1–34 (1990).
9. Kuo, P. H., Holland, H. D. & Danielson, A. *Geol. Soc. Am. Abstr. Progr.* **25**, 89 (1993).
10. Schopf, J. W. *Science* **260**, 640–646 (1993).

1. *The Guardian*, 2 August 1944; reprinted 2 August 1994.
2. Kelce, W. R. *et al.* *Nature* **375**, 581–585 (1995).
3. Sharpe, R. M. & Skakkebaek, N.-E. *Lancet* **341**, 1392–1395 (1993).
4. Toppari, J. *et al.* *Male Reproductive Health and Environmental Chemicals with Estrogenic Effects* (Report No. 290 for the Danish Environmental Protection Agency, 1–166, 1995).
5. Carlson, E., Giwercman, A., Keiding, N. & Skakkebaek, N.-E. *Br. med. J.* **305**, 609–612 (1992).
6. Adami, H. *et al.* *Int. J. Cancer* **59**, 33–38 (1994).
7. Bromwich, P., Cohen, J., Stewart, I. & Walker, A. *Br. med. J.* **309**, 19–22 (1994).
8. Auger, J., Kuntzmann, J. M., Czyglik, F. & Jouannet, P. *New Engl. J. Med.* **332**, 281–285 (1995).
9. Guille, L. J. *et al.* *Environ. Health Perspect.* **102**, 680–688 (1994).
10. Imperato-McGinley, J., Sanchez, R. S., Spencer, J. R., Yee, B. & Vaughan, E. D. *Endocrinology* **131**, 1149–1156 (1992).
11. PAHO (Pan American Health Organization) *Status of Malaria Programs in the Americas 40th Report* (World Health Organization, Washington DC, 1992).

Use of DDT has been banned or restricted in most Western countries for 20 or more years, largely because of its persistence (it has a half-life of about 100 years) and its accumulation and concentration in food chains and because of a range of deleterious effects on reproduction in wildlife. DDT is still detectable in us all (mainly in body fat), despite the ban on its use. It is not, however, this present burden of DDT that is the centre of concern; rather it is the level of exposure in the period from the 1940s to the 1960s when it was still in use. Could it have exerted effects then on developing males which have only become apparent in more recent years? To address this possibility we need to know the level of exposure of the general population to DDT (DDE) when it was in use and whether such exposure would be sufficient to cause reproductive defects.

When first introduced, DDT was used in ways that would be unacceptable now. For example, when Naples was liberated by the Allies in the war, the whole population was 'dusted' with DDT¹. Allied soldiers were similarly treated at recruitment and were issued with DDT-impregnated shirts when going overseas. In the tropics DDT was used in mosquito control and in many countries (including Britain) it was used in household fly-sprays. Today we would not contemplate using one of the safer modern pesticides in many of these ways, let alone a persistent chemical such as DDT. We can only guess what levels of exposure to DDT this led to in the general population, but measurements in dwellings in which DDT was used suggest that this exposure would probably be sufficient to exert significant anti-androgenic effects². Male alligators in Lake Apopka in Florida contain high levels of *p,p'*-DDE and have abnormally small penises, amongst other reproductive changes⁹. Administration of pharmaceutical anti-androgens (used in the treatment of prostatic cancer) to animals during fetal and neonatal life can result in a range of reproductive tract disorders (small penis, hypospadias, cryptorchidism)¹⁰, although sperm counts and testicular germ cell cancer in adulthood have not been studied.

If we are currently witnessing the legacy of DDT use in the past, can we at least comfort ourselves with the knowledge that things can only get better? Unfortunately not. Present use of DDT in developing countries (mainly for malarial control) probably exceeds the level of its use historically. Mexico and Brazil each used nearly 1,000 tonnes of DDT in 1992 according to World Health Organization figures¹¹. As well as posing a health threat in these countries, the export of DDT through animals, crops and the atmosphere to the countries where its use is now restricted remains likely. And we still have the matter of environmental oestrogens on our doorstep to clear up.

Perhaps the most remarkable aspect of the findings of Kelce *et al.*² is that it has taken 50 years to discover that the main metabolite of DDT is an anti-androgen. This, coming on top of the recent discovery of oestrogenicity of several abun-

Endothelin versatility

Katsutoshi Goto and Timothy D. Warner

THE peptides with the most potent activity in constricting blood vessels and elevating blood pressure in mammals are the naturally occurring endothelins (ET-1, ET-2 and ET-3). They also influence such activities as cell proliferation and hormone production and have been implicated in cardiovascular disorders ranging from hypertension to stroke and ischaemic heart disease¹. With such a store of actions it is not hard to understand why, since their discovery in 1988, the endothelins have stimulated great research activity in both academia and the pharmaceutical industry. These investigations have cleared much of the uncertainty surrounding the real roles of the endothelins, as discussed at a meeting in April*.

Dramatic advances have been made in three areas of endothelin research in particular: the breeding of 'knockout' mice, the characterization of the synthetic pathway of endothelin, and the development of potent non-peptide endothelin-receptor antagonists. The production of endothelin-deficient or endothelin-receptor-deficient mice by means of gene targeting has shown unexpectedly that the endothelins have crucial roles in normal embryonic development². ET-1-deficient mice for instance, have craniofacial and cardiac abnormalities at birth and die of respiratory failure soon after.

Surprisingly, the blood pressures of heterozygous mice, which have plasma and tissue levels of ET-1 half those of

dant environmental chemicals, must mean that our current methods of testing and surveillance of manufactured chemicals require modification so that the release of 'hormonally active' compounds to the environment can be minimized. For the present, we do not know whether such chemicals pose a significant risk to man, because we lack accurate data on the levels and routes of human exposure and on whether, and at what levels, adverse reproductive changes can be induced (in animals). Collecting this information is therefore a priority. Only then will we be in a position to assess whether these chemical skeletons in our cupboards are likely to have contributed to male reproductive disorders. □

Richard M. Sharpe is at the Medical Research Council Reproductive Biology Unit, 37 Chalmers Street, Edinburgh EH3 9EW, UK

wild-type mice, are elevated compared with their normal relations. After brief consideration one might draw from these observations the conclusion that ET-1 has a predominantly hypotensive effect within the body. This simple deduction may well be wrong, however, for heterozygous mice also exhibit an impaired chemoreceptor reflex and hence an abnormally high sympathetic tone (T. Kuwaki, Tokyo Univ.). Furthermore, adenovirus-mediated overexpression of ET-1 in rats prompts an increase in blood pressure (S. Télémaque, Univ. of Texas).

Additional evidence for the roles of the endothelins in the normal development of the embryo was provided by studies examining changes induced in mice by 'knockout' of the endothelin A (ET_A) receptor. These animals had craniofacial deformities identical to those in ET-1-deficient mice, indicating that the action of ET-1 on ET_A receptors is important in the development of neural-crest-derived tissues (K. Hosoda, Univ. of Texas). Interestingly, interaction of ET-3 with ET_B receptors also seems to be important for the development of neural-crest-derived melanocytes and enteric neurons. Mice deficient in these endothelin systems have a regional lack of epidermal melanocytes and develop megacolon, which are typical symptoms of Hirschsprung's disease³. The discovery that endothelins are intricately involved in neural crest development was entirely unpredictable from our previous knowledge but reminds us of the endothelins' far-ranging and potent effects. ►

*Fourth International Conference on Endothelin, London, UK, 23–26 April 1995.

Internal resistance

A CERTAIN bold space suit design has never been seriously developed. The conventional space suit fits loosely, and is pressurized with air. Imagine it being tightened until the pressure on the skin is exerted not by entrapped air, but by the fabric itself. Air pressure is then not needed. A fabric strong and inextensible enough to contain the pressure of the body would not even need to be gas-tight. A figure-hugging open weave would do. Between its restraining meshes, the skin would be directly exposed to space. It could sweat freely into 0% humidity, keeping the astronaut cool. He could walk through space in little more than an air-fed helmet and a body-stocking.

Daedalus has an earthly application for this idea: diagnostic electrical resistance tomography. By measuring the resistance between each point on the body's surface and every other point, it is possible to compute a resistance-map of its interior. The principle has been used to study the chest (the air-filled lungs are perfect insulators, and give a clear outline), and can clearly be taken further. Daedalus will fit the patient with a space body-stocking and immerse him, not in high vacuum, but in low-pressure neon. He can then serve as an electrode for a neon discharge display.

Current will enter the patient via a springy, mobile electrode. It will leave by a distributed glow discharge over his whole body, whose intensity at each point will indicate the local potential. High frequency a.c. will be used, which is not painful. The gas pressure and current density will be chosen to confine the neon glow to the patient's surface (as in those neon 'candles' whose discharges wander over their electrode surfaces). As the entry-electrode travels over his body surface, possibly in a helical scan, a TV system will digitize his changing pattern of personal illumination, and feed it into the system computer. In due course a three-dimensional whole-body resistance map should emerge.

Resistance tomography will reveal the body's interior in a novel neon light. Bones will show up as insulators, while nerves and blood-vessels will appear as conducting channels. Fractures, perforations and other injuries will shine out in high relief by the conductive paths they form through the damaged tissues. Furthermore, high-frequency currents can have a useful healing effect. Once the resistance-map has been computed, it will show where to attach electrodes later, so as to send such a current selectively along an internal damage-track, and encourage its rapid healing.

David Jones

Achieving a proper characterization of the pathways involved in endothelin biosynthesis has proved a more difficult goal than we might have predicted, absorbing for more than six years the efforts of a number of resourceful research groups⁴. The precursor of ET-1, big ET-1, may be cleaved from its own precursor by the action of furin at dibasic pairs of amino acids (R. Leduc, Univ. of Sherbrooke, Canada). ET-1 is then freed from big ET-1 by a specific endothelin-converting enzyme(s) (ECE). This latter enzyme has now been cloned and expressed and the molecular structures of a number of related ECEs derived. These have now been named ECE-1a, ECE-1b and ECE-2 and show different preferences for big ET-1, big ET-2 and big ET-3 as substrates (S. Kimura, Chiba Univ., Japan; N. Emoto, Univ. of Texas; O. Valdenaire, Hoffmann LaRoche, Switzerland).

Recent advances in the development of endothelin-receptor antagonists have accelerated the pace of investigations into the true pathophysiological roles of endothelins in mature animals and humans, and now hold out the prospect of the development of new therapeutic drugs. Potent peptide antagonists of the endothelin receptors have been available for the past five years, but these compounds have disadvantages. They are difficult to administer over long periods as they are not orally active, and they are particularly short-lived within the body. But despite their shortcomings, these compounds have been used to establish some interesting facts, such as the role of endothelin in monocrotaline-induced pulmonary hypertension in rats⁵.

Now, however, with the availability of orally active non-peptide antagonists, the active involvement of endogenous endothelins in a variety of pathologies such as myocardial ischaemia/reperfusion (M. Kojima, Takeda Chemical Industries, Japan; J. Pernow, Karolinska Inst., Stockholm), renal failure (D. P. Brooks, SmithKline Beecham, USA), experimental stroke (T. R. Patel, Univ. of Glasgow) cerebral focal ischaemia (F. C. Barone, SmithKline Beecham, USA), and some forms of systemic hypertension (E. Bird, Bristol-Myers Squibb, USA; J.-S. Li, Clinical Research Inst. of Montreal, Canada) has been confirmed. Interestingly, in the rat the non-peptide antagonist SB 209670 also reduces neointima formation after endothelin injury by balloon angioplasty (E. Ohlstein, SmithKline Beecham, USA). This once again draws our attention to the fact that we must consider endothelins not only as spasmogens in cardiovascular disease. They may well also be involved in other pathological vascular changes, such as the hyperplasia that occurs in hypertension and atherosclerosis.

The further development for which many researchers are still impatient is the production of non-peptide antagonists that are highly selective for the ET_A or ET_B receptor, together with inhibitors of the action of ECE. These would permit the fine delineation of the different roles these elements have in changes driven by the endothelin system. Such newer receptor antagonists may also help us finally to decide how many functional endothelin receptors there are. So far, only two genes for endothelin receptors have been identified in mammals, encoding the ET_A and ET_B receptor, respectively. However, assorted pharmacological evidence, including discrepancies between the rank order of affinity of antagonists to receptors and the effects of specific antagonists, speak to the pharmacologist of subclasses of endothelin receptors.

For the intracellular signalling events following endothelin receptor stimulation, it seems broadly that phosphatidylinositol hydrolysis and calcium mobilization are crucial in many short-term responses to the endothelins, such as contraction and secretion. Activation of mitogen-activated protein kinase, however, seems to be important for long-term events like cell division and hypertrophy (M. J. Dunn, Case Western Reserve Univ., USA).

What of the future? The continuing observations that endothelins are expressed in non-vascular tissues, such as the gastrointestinal and respiratory tracts, renal tissue, neural and glial cells (Y. Tomobe, Tsukuba Univ.; S. Uhlig, Univ. of Konstanz; F. E. Karet, Univ. of Cambridge; S. M. Cazaubon, ICGM, France) and that important crosstalk exists between the endothelins, cytokines (P. Klemm, William Harvey Research Inst., UK) and other vasoactive agents, all indicate areas that need to be further investigated. We are also less certain about the actions of endogenously produced ET-2 and ET-3 than we are of endogenously produced ET-1. Nonetheless, by using selective endothelin receptor antagonists, ECE inhibitors, and animals deficient in portions of the endothelin system, the physiological and pathological roles of all of the endothelins should eventually be revealed. □

Katsutoshi Goto is in the Department of Pharmacology, Institute of Basic Medical Sciences, The University of Tsukuba, Tsukuba, Ibaraki 305, Japan. Timothy D. Warner is at The William Harvey Research Institute, The Medical College of St Bartholomew's Hospital, London EC1M 6BQ, UK.

1. Rubanyi, G. M. & Polokoff, M. A. *Pharmac. Rev.* **46**, 325-415 (1994).
2. Kurihara, Y. *et al. Nature* **368**, 703-710 (1994).
3. Baynash, A. G. *et al. Cell* **79**, 1277-1285 (1994).
4. Xu, D. *et al. Cell* **78**, 473-485 (1994).
5. Miyauchi, T. *et al. Circ. Res.* **73**, 887-897 (1993).

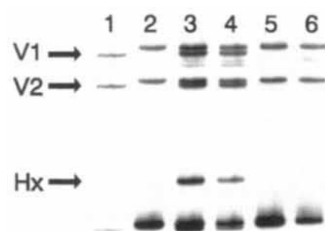


FIG. 2 Mutation analysis. Single-strand conformation polymorphism analysis of exon 11 of the *BRCA1* gene. Lane 1 contains the PCR product from the arrowed individual from pedigree 475, lane 3 is from her father and lane 4 from her mother. Lanes 2, 5 and 6 are normal controls. The two variant bands V1 and V2, and the heteroduplex band Hx, are arrowed. Methods are available from the authors.

As no wild-type sequence was present, we tested whether this individual was homozygous for this mutation, or was heterozygous for the AA₂₈₀₀ and had a deletion or polymorphism preventing amplification of the other allele. We screened lymphocyte DNA from her father and tumour DNA from her mother for the AA₂₈₀₀ mutation. Both her parents were heterozygotes for the AA₂₈₀₀ mutation, indicating that she is an AA₂₈₀₀ homozygote. We have also carried out haplotype analysis on this family and have found that the AA₂₈₀₀ homozygote is homozygous for five *BRCA1* flanking markers from *THRA1* to D17S579 (data not shown).

This woman is the first example, to our

knowledge, of a human homozygote for a high-penetrance cancer susceptibility gene. Her cells should be invaluable in studying the function of the *BRCA1* protein. It appears that her risk of cancer may be similar to that of heterozygotes for the AA₂₈₀₀ mutation, as she had breast cancer at age 32. The existence of viable *BRCA1* homozygotes has ramifications for the screening of high-risk women, as it is possible that a parent who is found to be heterozygous for a clinically meaningful mutation may have a second, unidentified alteration in the other *BRCA1* allele. The daughter of this parent could then be misdiagnosed as not having inherited a high-risk allele.

Marie Boyd

Fiona Harris

Robert McFarlane

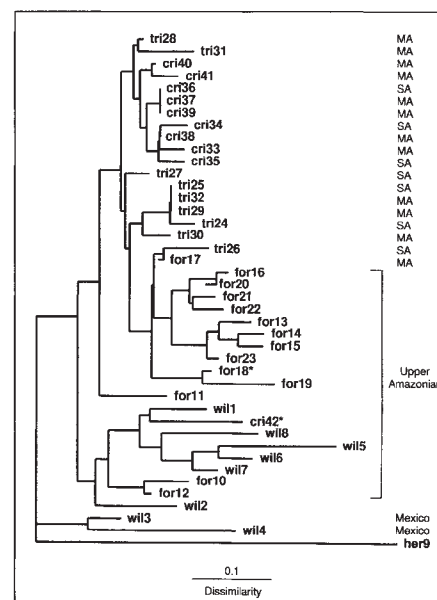
H. Rosemarie Davidson

Donald M. Black*

Cancer Molecular Genetics Group,
Beatson Institute for Cancer Research,
Cancer Research Campaign Beatson
Laboratories,
Bearsden, Glasgow G61 1BD, UK

*To whom correspondence should be addressed.

1. Ford, D. & Easton, D. F. *Br. J. Cancer* (in the press).
2. Easton, D. F. et al. *Am. J. hum. Genet.* **56**, 265–271 (1995).
3. Simard, J. et al. *Nature Genet.* **8**, 392–398 (1994).
4. Hall, J. M. et al. *Science* **250**, 1684–1689 (1990).
5. Miki, Y. et al. *Science* **266**, 66–71 (1994).
6. Futreal, P. A. et al. *Science* **266**, 120–122 (1994).
7. Castilla, L. H. et al. *Nature Genet.* **8**, 387–391 (1994).
8. Friedman, L. S. et al. *Nature Genet.* **8**, 399–404 (1994).
9. Shattuck-Eidens, D. et al. *J. Am. med. Ass.* **273**, 535–541 (1995).



Clustering of cacao accessions based on genetic dissimilarity and the neighbour-joining method. Illustrated is an unrooted, strict consensus tree of 6 tied trees (wil, wild; cri, Criollo; for, Forastero; tri, Trinitario). Right-hand side, geographical origin of accessions (MA, Mesoamerica; SA, South America). Two accessions, for18 and cri42, are in the Upper Amazonian cluster, yet their geographical origins are Mexico and Trinidad, respectively. The 'wild' accessions from Mexico are located at the base of the tree (wil3 and wil4).

Amazonian region, three cultivars (Criollo, Forastero, Trinitario), and the sister genus *Herrania*. We tested 60 decamer primers, of which 24 produced consistent, robust amplification products resulting in a total of 105 loci across all accessions⁵. Relationships among accessions were accessed by neighbour-joining⁹ on genetic dissimilarity (see figure). The analysis places the Mexican 'wild' collections at the base of the tree. Short branch lengths indicate a high degree of similarity among cultivars. Average dissimilarities between cultivars (see table) reveal fewer differences between Criollo and South American wild accessions than between Criollo and Mexican wild accessions. The basal position of Mexican 'wilds' and the greater similarity among South American wilds and all cultivars support the South American origin of present-day cultivated cacaos.

Our results indicate that Criollo cultivars are more similar genetically to South American germ plasms than to Mesoamerican 'wilds'. Other studies have

Origins of cacao cultivation

SIR — The origin and early domestication of cacao (*Theobroma cacao* ssp. *cacao* L.) remains an elusive issue as a result of uncertainty about the historical distribution of wild populations, ancient cultivations, human-mediated dispersion and continued crossing between cultivars and wild populations. Hypotheses on cacao origins either envisage independent domestication of the Criollo cultivar in Mesoamerica (from the ssp. *cacao*) and the Forastero cultivar in South America (from the ssp. *sphaerocarpum*)¹ or place the site of domestication of both cultivars in South America². Neither hypothesis has strong support from morphology or ecology. Historical evidence shows that cacao was grown in the Yucatan peninsula in

1566, when an early chronicler wrote: "They have sacred groves where they cultivated certain trees, like cacao"³. Discovery of plants similar to wild cacao from the Lacandon forest (*T. cacao* ssp. *cacao* f. *lacandonense* Cuatrecasas) in 'sacred groves'⁴ in Yucatan and additional collections from the Lacandon forest support a Mesoamerican origin for the Criollo variety, although recent genetic evidence supports an Amazonian centre of origin for both varieties^{5,6}. Using molecular markers to characterize new accessions, we now show that Mexican 'wild' collections are genetically distinct from the South American wild plants and modern cultivars. Recently found 'wild' Yucatan plants in Maya 'sacred groves' are most probably the closest living representatives of an ancient Maya cultivar.

Molecular markers are useful for determining relationships among wild and domesticated plants, especially random amplified polymorphic DNA (RAPD) markers in cacao^{5–7}. We surveyed 'wild' cacao populations from Mexico, wild populations from the Upper

GENETIC DISSIMILARITIES BETWEEN GROUPS OF CACAO ACCESSIONS

Group	Mean dissimilarity	s.e.m.
Mexican 'wilds' × S. Amer. wilds	0.111	0.006
Mexican 'wilds' × Forastero	0.111	0.005
Mexican 'wilds' × Criollo	0.104	0.008
S. Amer. wilds × Forastero	0.101	0.003
S. Amer. wilds × Criollo	0.082	0.003

1. Cuatrecasas, J. *Contr. U.S. natn. Herb.* **35**, 379–614 (1964).
2. Cheesman, E. E. *Trop. Agric.* **21**, 144–159 (1944).
3. de Landa, D. *Pap. Peabody Mus.* **18**, 164 (1941).
4. Gómez-Pompa, A., Flores, J. S. & Allphat Fernández, M. *Latin Am. Antiq.* **1**, 247–257 (1990).
5. Figueira, A., Janick, J., Levy, M. & Goldsbrough P. *J. Am. Soc. hort. Sci.* **119**, 1073–1082 (1994).
6. N'Goran, J. A. K., Laurent, V., Risterucci, A. M. & Lanaud, C. *Heredity* **73**, 589–597 (1994).
7. Russell, J. R., Hosein, F., Johnson, E., Waugh, R. & Powell, W. *Molec. Ecol.* **2**, 89–97 (1993).
8. de la Cruz, M. thesis, Univ. California, Riverside (1994).
9. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).

confirmed the close relationship of the Criollo and Forastero cultivars and South American wilds^{5,6}. These data imply that cacao cultivations of the Maya do not exist in modern germ-plasm collections. Managed groves were probably abandoned after European colonization, with a few remaining in isolation, protected by the Maya descendants. Rediscovered populations in Mexico are either derivatives of previously undetected wild cacao or remnants of ancient cultivars of the Maya people. The occurrence of the Yucatan 'wild' accession in known Maya 'sacred groves' supports the latter suggestion.

Marlene de la Cruz

Centro de Ecología,
UNAM Apdo, Postal 70-275,
CP 04510 Mexico City, Mexico

Richard Whitkus

Arturo Gómez-Pompa

Department of Botany and Plant Sciences,
University of California,
Riverside, California 92521, USA

Luis Mota-Bravo

Department of Ecology and Evolutionary
Biology,
University of California,
Irvine, California 92717, USA

Tuatara sex determination

SIR — Temperature-dependent sex determination (TSD) exists in three orders of living reptiles¹⁻³. Here we report evidence that both surviving species of a fourth order (Sphenodontida) also exhibit this phenomenon. This finding is important to the conservation of sphenodontids, and increases confidence in the suggestion that TSD is the ancestral sex-determining mechanism in reptiles⁴.

We suspected that the lizard-like reptile tuatara (*Sphenodon punctatus*) exhibited TSD after observing total sex biases among locatable offspring for three zoo-incubated clutches. We subsequently sexed juvenile *S. punctatus* remaining in New Zealand from experiments in 1986 in which wild eggs were incubated under controlled conditions^{5,6}. Live offspring were sexed by laparoscopy and available carcasses by dissection and/or histology. Incubation temperature had a highly significant effect on sex ($P=0.001$). Sex ratios (F:M) were 9:0 at 18 °C, 31:3 at 20 °C and 4:13 at 22 °C, temperatures within the range experienced by natural nests⁷. Although 32.4% of eggs failed and were unsexed, differential mortality of sexes pre-hatching is, as in other reptiles⁸, unlikely to explain the results. At 20 °C, for instance, 83/115 eggs hatched and among 34 that hatched 91.2% were female. If we assume this sex ratio for all 83 hatchlings and also assume the extreme case of all dead embryos being male, the

calculated sex ratio (76 F: 39 M) is still significantly biased ($P<0.001$). Differential mortality post-hatching is unlikely, as sex ratios at 20 °C show a female bias both in juveniles sexed live (23:2; $P<0.001$) and in those sexed dead (8:1; $P=0.04$).

Data from an egg-incubation programme generating *S. guntheri* for reintroduction to the wild⁹ also suggest that TSD exists. Wild eggs were incubated half-buried at -430 kPa and at 18 or 22 °C or a variable temperature regime (mean temperature 19.8–20.6 °C; temperature varied gradually over 18→23→18 °C during incubation). Overall hatching success was 88.0%. Sex ratios (F:M) among dead juveniles (live offspring were too small for laparoscopy) were 17:0 (18 °C, $P<0.001$), 3:0 (22 °C, $P=0.250$) and 0:7 (variable, $P<0.02$). The male bias under the variable regime may result from temperatures >22 °C for at least 2 weeks in the middle of incubation, which in other reptiles is within the temperature-sensitive period¹. Both species exhibited 100% females at 18 °C and a tendency towards males under a warmer regime(s). These observations fit the type Ib² or FM^{1,10} pattern of TSD in reptiles, though further studies are necessary to rule out a type II (FMF) pattern³.

In rare sea turtles, some incubation practices performed without knowledge of TSD have apparently resulted in male-biased sex ratios of reduced benefit to conservation^{11,12}. Tuatara are restricted to 30 offshore islands of New Zealand¹³. Fortunately, the range of temperatures used in incubation programmes for both species has resulted in an estimated 68–77% of offspring being female, a favourable ratio for conservation purposes.

Alison Cree

Department of Zoology,
University of Otago,
Box 56, Dunedin, New Zealand

Michael B. Thompson

Department of Zoology (A08),
School of Biological Sciences,
University of Sydney,
New South Wales 2006, Australia

Charles H. Daugherty

School of Biological Sciences,
Victoria University of Wellington,
Box 600, Wellington, New Zealand

- Lang, J. & Andrews, H. V. *J. exp. Zool.* **270**, 28–44 (1994).
- Ewert, M. A. & Nelson, C. E. *Copeia* **1991**, 50–69 (1991).
- Viets, B. E. *et al. J. exp. Zool.* **270**, 45–56 (1994).
- Janzen, F. J. & Paukstis, G. L. *Q. Rev. Biol.* **66**, 149–179 (1991).
- Thompson, M. B. *Resp. Physiol.* **76**, 243–256 (1989).
- Thompson, M. B. *J. Zool.* **222**, 303–318 (1990).
- Thompson, M. B., Packard, G. C., Packard, M. J. & Rose, B. J. *Zool.* (in the press).
- Hutton, J. M. *J. Zool.* **211**, 143–155 (1987).
- Cree, A. *et al. Contr. Herpet.* **11**, 377–385 (1994).
- Ewert, M. A., Jackson, D. R. & Nelson, C. E. *J. exp. Zool.* **270**, 3–15 (1994).
- Mrosovsky, N. *Biol. Conserv.* **23**, 309–314 (1982).
- Wibbels, T. R. *et al. Proc. 1st int. Symp. Kemp's Ridley Sea Turtle Biology, Conservation and Management* Galveston, Texas (eds Caillouet, C. W. & Landry, A.) 77–81 (Sea Grant College Publ., Galveston, Texas, 1989).
- Daugherty, C. H., Cree, A., Hay, J. M. & Thompson, M. B. *Nature* **347**, 177–179 (1990).

Mouse knockouts rule OK

SIR — The value of genetic analysis to the study of learning and memory, specifically the value of mouse null mutants, has recently been questioned by A. Routtenberg in Scientific Correspondence (*Nature* **374**, 314–315; 1995). The objection was raised that mice harbouring null mutations are "reactionisms" (*sic*): another gene product, or system of gene products, may compensate for the missing gene product. As a consequence, data generated in the analysis of null mutants may be uninterpretable, presumably because of the complex interactions among the many genes involved in learning and memory.

The objection is based on the ill-founded assumption that an inability to observe a phenotype must mean that there is no phenotype to observe. But the maintenance of a gene is *prima facie* evidence that its absence would occasion a decrease in fitness (a phenotype), though not necessarily one easily observed under limited laboratory conditions. This mistake was compounded by overstating the significance of the synergism between mutations that frequently produces new phenotypes: this is not a novelty of mouse genetics, but a common occurrence even in *Escherichia coli*.

Complex biological processes will have complex genetics. This truism can easily be illustrated by considering mutations in *E. coli*: mutations affecting the catabolism of lactose define a relatively simple genetic pathway, whereas those affecting the doubling time of *E. coli* on rich medium would define an extremely complicated genetic system. Only in such simple pathway systems as intermediary metabolism can we expect simple genetic pathways; as learning and memory in mammals are among the most complex of biological processes, we can only expect their genetics to be complex.

We should not abandon genetic analysis because of its unwelcome message that learning and memory are more complicated than our heuristic theories allow. Nor should we hope that genetics or some future method of analysis will effect a miraculous simplification of these processes. Rather we should make use of, and integrate, data from all available methods. In the interplay between the production of mutants and their phenotypic analyses, our ability to generate mouse mutants has clearly outstripped our ability to analyse their phenotypes. We are especially in need of more refined methods for studying neural processes in mice.

U. Hochgeschwender

M. B. Brennan

Unit on Molecular Genetics,
National Institute of Mental Health,
NIH, Bethesda, Maryland 20892, USA

Native Al and Si formation

SIR — The natural occurrence of elements such as aluminium and silicon in their native form is usually ascribed to either contamination or the presence of strongly reducing fluids¹ (for example, pure methane or hydrogen, although such fluids have never been found in nature). Here we describe the formation of crystals of native aluminium, silicon, titanium, iron and platinum inside the quartz tubes that we used to sample hot volcanic gas jets associated with Kudriavsky volcano on Iturup island in the Kuril arc.

The gas emitted by the volcano consists mainly of water (93–97 mol%), CO₂ (1.3–4.4 mol%), H₂S (≤ 1.25 mol%) and SO₂ (≤ 1 mol%), with small quantities of H₂, HCl and HF (ref. 2). It is important to note that the hydrogen concentration does not exceed 0.65 mol%, and the methane concentration is two orders of magnitude less. We determined the oxygen fugacity, f_{O_2} , by two methods: thermodynamic calculations based on the CO/CO₂ and H₂S/SO₂ of the gas, and direct measurement with an electrochemical cell³. All measured and calculated values lie between the Ni–NiO₂ and haematite–magnetite buffers, indicating an oxidizing environment.

Fumarole sublimates were sampled using concentric quartz tubes inserted into cleaned volcanic conduits. The

upper half of each tube protrudes from the conduit and is air cooled. The two fumaroles examined by this method had gas temperatures of 930 and 705 °C — the corresponding temperature gradients in the sampling tubes were approximately 150 and 300 °C, respectively. The tubes were removed from the fumaroles after two months for laboratory examination.

In the tube from the 930 °C fumarole, we identified precipitates of aegirine, fayalite, augite, rutile, Mg-calcite, apatite, magnetite, haematite, pyrite, pyrrhotite, molybdenite, sphalerite, galena, cubanite and complex sulphides. At the hot end of the tube, thin plates of platinum and iron were found. The Pt contained a small admixture of iron (≤ 0.3%) whereas the iron was pure and homogeneous.

The tube from the 705 °C fumarole was completely filled by sublimates having a total mass of 66 g and containing 78% sulphur, 12% halides (mainly NaCl and KCl) and 10% water-insoluble components. Halite and sylvite precipitated mainly in the hot part of the tube, whereas sulphur precipitates were found only in the upper part of the tube. Several minerals were identified, including SiO₂, pyrite, plagioclase, K-feldspar, clinopyroxene, aegirine, augite and radite garnet, molybdenite, avogadrite, alpaolite and complex sulphides. Associated with the minerals, we found particles of native aluminium (96.11% pure by mass) and silicon (100% pure) and one plate of titanium (99.8% pure).

The aluminium particles were found in the centres of Cl-rich spheres, growing outward to the rim as a sponge or honeycomb structure (*a* in the figure) and ranging in size from 0.1 to 3 mm. We could not determine the precise composition of the material in the sponge itself, but it appears to be enriched in Cl (as AlCl₃(OH)_n). The outermost parts of the Al are partly replaced by KAl(SO₄)₂·nH₂O. The native silicon occurs both as thread- and diamond-shaped crystals within the Al, and as separate crystals (~0.1 mm in length) within Cl-rich spheres. The figure (*b*) shows one particle in which a large Si crystal is surrounded by Al–Si alloy with a texture typical of a eutectic.

Our system of sample preparation makes it unlikely that the native elements represent contamination after retrieval of the sampling tubes. The texture of the particles also argues that they formed *in situ*.

First, we note that the metals (Al, Si, Ti, Pt and Fe) formed

only in zones where chlorides and fluorides precipitated and, under the sampling conditions, the halides will have remained liquid until the tubes were removed from the fumarole. We propose that the massive precipitation of halides plays a key role in the formation process of these metals. Geochemical modelling with the IVTANTERMO database⁴ implies that in the volcanic gas the main Si-containing species at temperatures >700 °C would be SiO. Low-valence compounds of Al have also been described^{5,6} (such as AlCl). Native Si and Al might thus have formed as a result of disproportionation reactions, for example 2SiO→Si+SiO₂ or 3AlCl→2Al+AlCl₃ (ref. 6). The salt melt that condenses in the sampling tubes (and presumably on the inner surface of the volcanic conduits) may then trap unstable compounds such as AlCl and SiO, and provide an oxygen-free medium in which crystals of the native metals can grow. Intriguingly, particles of native aluminium have also been found associated with pure rhenium sulphide sublimates⁷ at the same volcano (C. Stanley, personal communication).

To test this hypothesis, we conducted a pilot experiment in which a salt trap was placed in the jet of an 800 °C fumarole. The mixture of Si and SiO₂ that precipitated in the trap had a Si:O ratio near 1:1, suggesting a formation mechanism involving SiO. We cannot discount the possibility that conditions in the sampling tubes — in particular, disequilibrium caused by the blocking of the tube by the sublimates — also played a role. But we argue that similar precipitates of native elements could form (at least transiently) in other hydrothermal systems in which the temperature is high enough to stabilize a halide melt. The high solubility of halides in hydrothermal fluids suggests, however, that a record of this unusual process is unlikely to be preserved for a significant length in time.

M. A. Korzhinsky

S. I. Tkachenko

K. I. Shmulovich

*Institute of Experimental Mineralogy,
Russian Academy of Sciences,
Chernogolovka, 142432 Moscow, Russia*

G. S. Steinberg

*Institute of Volcanology and Geodynamics,
Academy of Natural Sciences of
Russian Federation,
Sakhalin, Russia*

1. Ryabchikov, I. D. & Novgorodova, M. I. *Dokl. Akad. Nauk SSR* **258**, No. 6 (1981).

2. Tkachenko, S. I. et al. *Dokl. Akad. Nauk SSR* **325**, 823–828 (1992).

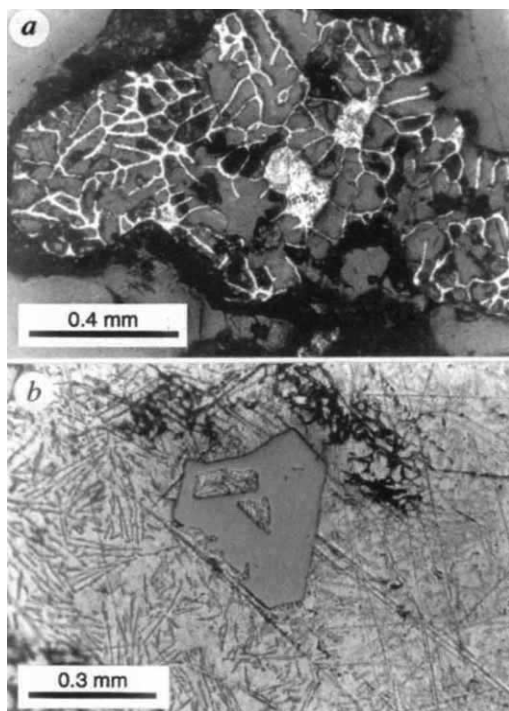
3. Rosen, E., Osadchii, E. & Tkachenko, S. *Chem. Erde* **53**, 219–226 (1993).

4. Glushko, V. P. (ed.) *Thermodynamic Properties of Individual Substances* 4 vols (Nayka, Moscow, 1978–81).

5. Schafer, H. *Chem. Transp. React.* (Academic, New York, 1964).

6. Osadchii, E. G., Alekhin, Y. V. & Kuzenko, V. I. *Chem. Erde* **45**, 331–337 (1986).

7. Korzhinsky, M. A. et al. *Nature* **369**, 51–52 (1994).



Native metals from the 705 °C tube. *a*, Al-sponge with two Al droplets which diminished in size during polishing. *b*, Particle of Al-Si alloy with eutectic structure and large Si crystals.

Rousing the sleeping genes

David Weatherall

Altered Fates: Gene Therapy and the Re-tooling of Human Life. By Jeff Lyon and Peter Gorner. Norton: 1995. Pp. 636. \$27.50.

WHETHER or not gene therapy turns out to be "a medical revolution unparalleled in human history", as the jacket of *Altered Fates* predicts, few fields of medical research have received so much advance publicity during their gestation. Already two journals are devoted to it, several accounts of its history have been written and it features in numerous popular books on the potentials and dangers of the new genetics. And now the trials and tribulations of its early days, together with what it may hold in store for the future, are described in more than 600 pages of unrestrained enthusiasm.

No one could deny that the field is busy. By the end of 1994, at least 100 clinical protocols involving gene transfer had been approved and it was estimated that the number of patients being treated, in the region of 300, had doubled since 1993. But although, touch wood, this novel form of therapy seems to have been unexpectedly free of side effects, there is very limited evidence of its clinical efficacy: the symptoms of two children with an immunodeficiency syndrome resulting from adenine deaminase (ADA) deficiency appear to have been controlled successfully with repeated injections of lymphocytes into which normal ADA genes had been inserted; a patient with familial hypercholesterolaemia has shown a modest fall in blood cholesterol level after gene replacement; and a few cancer patients have responded to different forms of genetic manipulation.

Yet even here there are uncertainties. Despite the fact that the ADA and related replacement experiments have led to a successful patent application covering all *ex vivo* gene therapy in the United States, details of the work are still to be published. There are uncertainties about the interpretation of the results reported for the treatment of hypercholesterolaemia, and little is known about the outcome of the cancer patients.

So although it is a reasonable guess that gene therapy will have much to offer medical practice in the long term, a great deal more research is required, both into better ways of delivering genes to different cell populations and to ensure they work adequately in their new homes. What is more, it is still not clear whether some of the vectors used for gene transfer are absolutely safe.

Writing a popular book about research still in its infancy presents formidable problems. In particular, in trying to

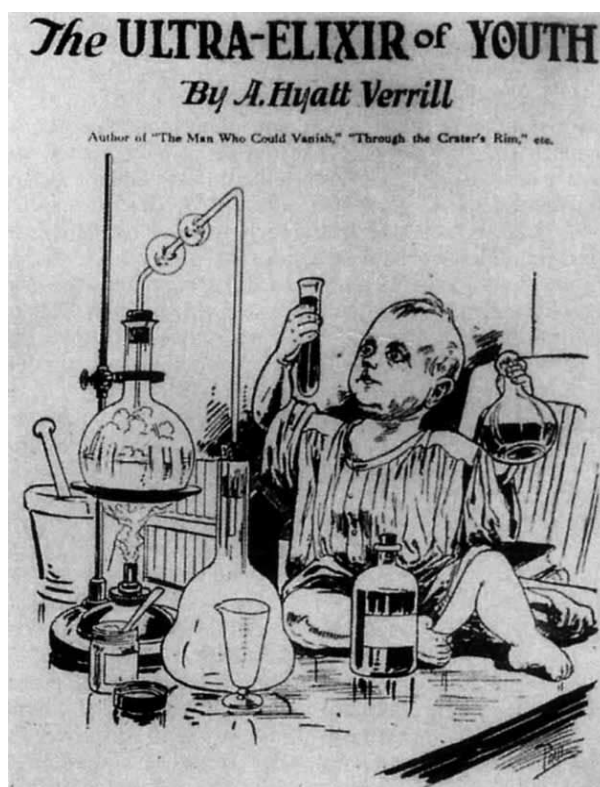
convey its excitement and potential it is tempting to play down the 'ifs' and 'buts', making it difficult to present a genuinely balanced account of the current state of the art. And if the authors are not scientists working close to the field and able to assess its literature, they must rely on the objectivity of the researchers that they interview. So what do two Pulitzer prizewinning journalists make of gene therapy after several years spent in the company of its practitioners?

The first half of *Altered Fates* is devoted mainly to the ADA story and French Anderson and his colleagues at the US National Institutes of Health (NIH) who began this research. After providing a particularly vivid pen-sketch of Anderson, the authors tell of his long struggle to obtain permission to carry out this experiment, the scepticism of his scientific colleagues about its feasibility or even whether it should have been attempted at all, and the personal lives of the young patients and their families. In the end, the authors come out strongly on the side of the NIH team, describing the ADA experiment as a "scientific milestone of incalculable

magnitude". This may well be true, but surely now that this and related work on cancer at the NIH has led to a successful patent application, the details must be published so that the scientific community can judge for itself.

Parts of the second half of the book go over ground popularized elsewhere, particularly the discovery of the molecular pathology of single gene disorders such as Duchenne muscular dystrophy and cystic fibrosis. But it also explores the possibilities for their treatment by gene therapy and assesses the role of genetic manipulation for the management of some of the common disorders that plague Western societies such as heart disease, cancer and dementia. It finishes with a lively section on the application of genetic manipulation to the prolongation of life.

Although much of the book deals with areas of research that are now well established and in which the technology is based on the well-tried tools of molecular and cell biology, the final chapter, entitled "The Ultimate Frontier", moves into the uncharted new field of biogerontology. Here, we are told, gene therapy will be used to "rouse the sleeping genes in ageing people that once kept them young, lean, and muscular". No longer young and never having been lean or muscular, I read on at a pace usually reserved for the closing pages of a novel by P. D. James. Disappointingly however, it seems I am destined to remain geriatric and flabby for a while yet. Worse still, it is not always



BILLING itself as "a picturesque tale of the biological possibilities in the field of modern science", A. Hyatt Verrill's story *The Ultra-Elixir of Youth* takes as its subject the discovery of "a gas which will prove the Elixir of Life". It fictionalizes the process of developmental reversal through endocrine treatment discussed in a 1922 essay by Julian Huxley. The story yokes embryos with test tubes. The picture appears in *Babies in Bottles: Twentieth-Century Visions of Reproductive Technology* by Susan Merrill Squier. Rutgers University Press, \$48 (hbk), \$17 (pbk).

clear how some of the experiments in biogerontology described by the authors are related to human physiology, let alone to gene therapy. It may come as a surprise to endocrinologists, as it did to me when reading about the anti-cancer and anti-ageing properties of dehydroepiandrosterone (DHEA), to learn that "it is a build-up hormone that counters the tear-down fight-or-flight hormone cortisone. . . cortisone virtually pours from the thymus gland when mice are placed in wheel-like drums and spun round and round, but DHEA nullifies its effect". The scientist who is carrying out this work is quoted as saying: "the more you get into DHEA, the more it sounds like snake oil. . . it's working here, it's working there. Omigod!". (I was unsure about the derivation of the last word, but in this context it sounded appropriate enough when I read it aloud.)

Much of *Altered Fates* is couched in this way. Overall, it does not try to interpret the scientific complexities of gene therapy for a nonspecialist audience, but rather it highlights its considerable medical possibilities. Although it paints an extremely rosy picture and tends to play down many of the difficulties and disappointments that undoubtedly lie ahead, I suspect this is not the fault of the authors; clearly, much of what is said is taken directly from transcripts of the numerous interviews with the scientists they encountered during their background research. For this reason their story offers a fascinating commentary on the complex sociology of the gene-therapy field and, one suspects, of much of current research.

Nonspecialist readers who can live with this rather extravagant style of writing will learn a great deal about the background to gene therapy and gain some vivid insights into the way in which the biomedical sciences work, particularly in the United States. For the most part it is a lively piece of thoroughly researched journalism, which, although it says little about the underlying science, paints an enlightening picture of the lives, motivations and aspirations of today's researchers. Some readers may wonder if the larger-than-life characters depicted could possibly exist; from personal knowledge I can assure them that they do. □

David Weatherall is at the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

New in paperback

Concepts of Modern Mathematics by Ian Stewart. Dover, \$8.95. A humorous and anecdotal introduction first published in 1975.

Ancient North America: The Archaeology of a Continent by Brian M. Fagan (2nd edn). Thames and Hudson, £16.95. An account for both general readers and students.

Going places

Desmond King-Hele

A Travel Guide to Scientific Sites of the British Isles. By Charles Tanford and Jacqueline Reynolds. Wiley: 1995. Pp. 344. £12.99 (pbk).

SCIENCE, as much as literature, has a sense of place. Many scientific advances are linked with particular sites, and all scientists are born somewhere. In Britain the literary community has been far more successful in exploiting the benefits of local interest: Stratford makes much of William

not always relevant or scientific, followed by chatty and useful travel hints, with opening hours and appropriate telephone numbers for museums, churches and so on. After the main guide come maps and indexes of names, places and subjects.

The first region is London, where we are offered a good selection of sites. But a much fuller guide is already available, in *London Science* by Dennis and Sylvia Rosen (Prion, 1994). Curiously enough, the memorials in Westminster Abbey and the blue plaques on London buildings are not listed in either book. Perhaps that would be too systematic and boring; but a reference to Alan



Home truths — Down House, the birthplace of Darwin's theory of evolution by natural selection.

Shakespeare; Shrewsbury makes little of Charles Darwin. There are many geographical guides to literary landmarks, few to sites with a scientific aroma.

So this new guide is much to be welcomed. The authors have visited most cities in the British Isles, and a wide spectrum of towns and villages, in pursuit of scientific anecdotes and ambience. Perhaps a famous or obscure scientist was born or worked there; perhaps the town's museum is of scientific interest. This procedure may seem rather hit-or-miss, but it makes for easier reading than a pedantically comprehensive listing.

The book begins with 48 pages about notable advances in science made in Britain, from Stonehenge to DNA. This lively sketch is refreshingly free of the fetters of political correctness that bind the chain-gang of academic historians, who may wince at some of the more simplistic judgements. The selection of topics is arguable, too, but the resulting concoction is pleasantly readable.

The main guide occupies the bulk of the book. The authors divide the British Isles into 11 regions, and within each the chosen towns are listed alphabetically. The entry for a typical town comprises a short essay or story, easy to read even if

Symons's *Behind the Blue Plaques of London* would be helpful.

After London we visit six regions of England — southeast, southwest and so on — then Wales, Scotland (mainland and island) and Ireland. The choice of towns, seemingly capricious, is wide enough to uncover a rich tapestry of little-known facts and stories from the scientific past. We meet William of Occam, appropriately clean-shaven, in a stained-glass window at Ockham Church near Guildford. We admire the young curate Jeremiah Horrocks at Much Hoole near Preston as he observes the transit of Venus across the Sun in 1639, during intervals between the Sunday services. We are urged to make pilgrimages to the monuments of palaeoscience, the stone circles in the Orkney Islands and at Callanish near Stornoway. We are told of Benjamin Franklin pouring oil on the wavelets of Derwentwater; and of Manchester producing thousands of mourners at the funeral of John Dalton (an honour reserved today, it seems, for notorious criminals or politicians).

The book is a sitting target for petty criticism because it is so selective and often drifts away from science. Six Cambridge colleges are covered and the rest

are ignored; Liverpool has only a few lines, Manchester nine pages; Portsmouth and Southampton are not mentioned; there is little of scientific relevance in the entries on Edgehill, Inverness and Wrexham among others. Inevitably, too, the bland and sweeping career-summaries fall into error at times. One amusing mistake is for Ashbourne in Derbyshire, where the entry confidently begins: "Erasmus Darwin was partial to the bottle". In fact, he was famous as a doctor for having "sobered the county of Derby", as Anna Seward put it. (Perhaps it was bottled water he was partial to?)

Although all these criticisms are valid, I liked the book. It is lively, positive and reasonably accurate. Above all, you feel that the authors have actually been there. The book abounds in strange stories, well known and little known, and I confess that I went on reading it to the end — the ultimate tribute to a guide book from a reviewer. □

Desmond King-Hele is editor of Notes and Records of the Royal Society.

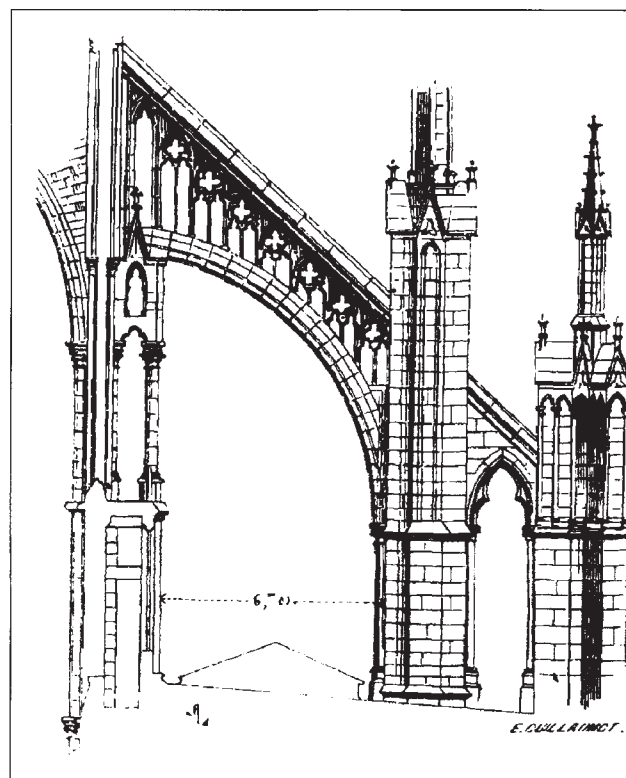
What mad pursuit

Stuart Sutherland

Uncovering Lives: The Uneasy Alliance of Biography and Psychology. By Alan C. Elms. Oxford University Press: 1995. Pp. 315. £18.99, \$25.

The Price of Greatness: Resolving the Creativity and Madness Controversy. By Arnold M. Ludwig. Guilford: 1995. Pp. 302. \$26.95.

PERHAPS the most disreputable branch of psychology is 'personality theory', for it is a theory that exists in name only. Despite the repeated application of factor analysis to aspects of personality, nobody can agree on what constitutes its basic traits. Nor is there even agreement that there are basic traits, because there is evidence that people act in different ways in different situations — they may be aggressive in one, timid in another; sometimes warm, sometimes cold; mean on one occasion, generous on others. The only interesting and imaginative contribution to the study of personality was made by Sigmund Freud. Unfortunately, it has no empirical validity: Freud, as he admits, talked his patients into believing what he wanted them to believe, as do his successors, including those therapists who falsely persuade their patients that they have suffered child abuse. Although, mercifully, the practice of psychoanalysis is declining in both the United Kingdom and the United States and the theory is largely ignored by psychologists, it still attracts literary magpies who use it in attempts to



NINETEENTH-century illustration of the buttressing system at Amiens cathedral. Taken from *The Stone Skeleton: Structural Engineering of Masonry Architecture* by Jacques Heyman, which unravels the science behind piers, pinnacles, towers, vaults and domes. Cambridge University Press, £35, \$59.95.

uncover the hidden meaning of texts. It is easy to see why Freud's theory is so alluring: it is based on the sex drive, and its main actors — the id, ego and superego — battle with one another in ways worthy of the Greek gods. Perhaps of most importance, no interpretation based on psychoanalysis can ever be refuted. Anything goes.

Alan Elms has chosen to apply the psychology of personality to uncovering the hidden motives and emotional influences that shaped the lives and writings of well known figures. He begins with some pious and extremely detailed advice to would-be psychobiographers (of whom the first was Freud in his notoriously inaccurate study of Leonardo da Vinci). He recommends them to focus on one aspect of their subject's life; to investigate archives, not forgetting to be nice to their curators; to interview surviving friends and relatives; to record the interviews, first making sure the tape recorder is working; and to concentrate on the meaning of contradictions in the subject's actions and writing.

He claims that modern personality theory should be used as well as psychoanalytic ideas, while taking the truth of both for granted. Almost his only nonpsychodynamic interpretation is that General Douglas Haig had an "authoritarian personality" — that is, obedience to one's superiors and insistence on obedience from inferiors, prejudice against outgroups, blaming others for one's own mistakes and so on. Unfortunately, the authoritarian personality has long since

been discredited: the traits alleged to compose it simply do not go together.

Elms's free use of psychoanalysis can be illustrated by his accounts of two of his subjects. Because Isaac Asimov gave a penetrating description of agoraphobia in one of his novels, Elms thought Asimov might be agoraphobic himself and had the cheek to ask him if that were so. Asimov replied that he was not an agoraphobic, but Elms took this and further denials as evidence that Asimov must be one, concluding that he wrote science fiction to escape from himself. On the other hand, in his analysis of B. F. Skinner's motives for writing *Walden Two*, he put it to Skinner that it was written for therapeutic reasons to extricate himself from the midlife crisis (a concept for which there is no evidence). Skinner, possibly in order to stop Elms pestering him, civilly agreed that this was partly the reason. Elms then goes on to trace the ways in which the contents of *Walden Two* relieved the worries Skinner had at the time: for example, in his Utopia, children are raised by professionals, a practice that would have spared Skinner the trouble of rearing his own offspring. Psychodynamic hypotheses are confirmed both by their denial and by their acceptance.

Uncovering Lives applies the same treatment to many others, including Carl Jung and the former US president George Bush. Elms maintains that Bush began the Gulf War "to prove his masculinity to himself": he seems unaware that Bush may have had other reasons for taking this decision. Elms writes well and

he stalks his prey in the unconscious minds of his subjects with diligence and gusto. *Uncovering Lives* should appeal to anyone who likes fiction, particularly detective stories.

In *The Price of Greatness*, Arnold Ludwig, a psychiatrist, accepts one of Elms's dicta by focusing on one feature of personality, creativity, but he takes a very different tack. Instead of examining the individual, he has undertaken an epidemiological study of 1,004 eminent Westerners who lived in this century. His sample was based on biographies of eminent people written between 1960 and 1990 excluding, one is thankful to report, famous criminals. Even so, there is a big difference between creativity and eminence, which Ludwig does not sufficiently acknowledge. He includes Marilyn Monroe, whose creativity is surely open to doubt, whereas Winston Churchill (also included), although creative, might not have become eminent had it not been for the Second World War. Moreover, luck plays a much greater part in eminence than Ludwig allows: had Roosevelt not died when he did, Truman might never have been heard of. "Full many a flower is born to blush unseen". Ludwig's study is marred by the lack of a control group of ordinary people with whom the backgrounds and characteristics of the great could be compared: in consequence he is often limited to comparing the characteristics of people eminent in different fields.

His most secure finding concerns the lifetime prevalence of mental illness in the great. Although overall it is little higher than in the population as a whole, there are striking differences between the different professions. Seventy per cent of poets had experienced depression and almost as many novelists and artists, while the prevalence of mania in these groups was about 15 times higher than in the general population. This will come as no surprise to readers of Kay Jamison's excellent book *Touched with Fire* (1993). Readers of *Nature*, however, can be reassured: eminent scientists have an exceptionally low risk of mental illness. It remains an open question whether creative authors write to ease mental conflicts or whether they become eminent because these conflicts enrich their work.

For the rest, there are few surprises in Ludwig's sample of the great. Firstborns are overrepresented, but then it has been known for some time that they tend to be more intelligent than others; writers and artists are more likely to be eccentric than are businessmen; homosexuality is common in writers. Of more interest, scientists and writers have one trait in common — both tend to have been sickly or poorly adjusted when young.

Ludwig ends by allocating scores for creativity to each of his subjects. He lists

30 predictors that discriminate between those in the top and bottom quartiles and uses a regression analysis to predict in which quartile subjects lie. Surprisingly, 95 per cent of subjects were correctly placed, but because three times as many men as women were in the top quartile he could have scored 75 per cent on this criterion alone.

Moreover, he was not predicting, he was retrodicting from data already available. How far the 30 variables would predict degrees of creativity if they were applied to a different population of eminent people we do not know. Nor, despite the 55 pages of tabular data set out in an appendix, do we know the true causes of greatness. Talent and perseverance are obviously important, as is luck.

The two books contrast sharply with one another. One focuses entirely on individuals, the other provides a statistical analysis of a large population. One uses theories too vague to substantiate (and sometimes already discredited), the other uses a sledgehammer approach that in the absence of any hypothesis — let alone theory — leaves the reader floundering. Nevertheless, they have two things in common. They both fail to take into account the influence of heredity on human characteristics (the evidence for which is increasingly strong) and they both read like parodies of their respective methodologies. □

Stuart Sutherland is at the Laboratory of Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Much about time in little space

Peter Landsberg

About Time: Einstein's Unfinished Revolution. By Paul Davies. Simon Schuster/Viking: 1995. Pp. 316. \$24, £17.

ALMOST every possible aspect of 'time' and its associated scientific and philosophical problems are touched on in this book. An obvious omission is perhaps the question of a multidimensional time. But apart from that, every problem one might expect, as well as several one might not, is raised: cyclic time, the twin paradox of special relativity, black holes, worm holes, many-universe theories, limits to computing, the tunnel effect, naked singularities, the anthropic principle, imaginary time, advanced and retarded potentials in electromagnetism; there are even brief nods in the direction of Nietzsche, Kant and Bayes's rule of probability theory. Einstein's name is milked for all it is worth; it is used more than necessary, and there is even a one-page account of his life at the end.

So what are we to make of this, Davies's nineteenth popular science book? It must be said at once that he makes a great, and on the whole successful, effort to explain all the notions and ideas he introduces, however complicated; this is no mean achievement as he eschews all equations. Still, it struck me as a pity that the first space-time diagram appears only 17 pages after the problem of the twins has been extensively discussed. This part might have been abbreviated. In fact, quite often while reading the book I felt that it would have profited from briefer and sharper explanations.

As already pointed out, Davies covers a huge amount of material, and his references span the period from Lucretius and Plato via Eddington (1929), Tolman

(1934) J. D. North (1965) to recent papers in *Physical Review* (1991, 1992), including Frank Tipler's controversial 1994 book, whose title (*The Physics of Immortality*) ought to have made it a bestseller (for a review see *Nature* 371, 115; 1994). To my considerable surprise, Davies even resurrects Herbert Dingle's *Science at the Crossroads* (1972), a book whose life expectancy never struck me as more than a few years. The book makes much use of the 'timewarp factor'. The budding physicist will be relieved to discover from the excellent index that this is nothing more mysterious than the unusual relativistic time dilatation.

Davies's book can be read easily and most readers will find in it matters of absorbing interest, so I can recommend it fairly strongly. Apart from the individual scientific aspects noted above, there are of course certain rather basic problems. Is time travel a phantasy? Can tachyons be ruled out? Is time an illusion of the human mind? What is responsible for the directedness of 'time'? Here the trouble is that there seem to be as yet no real answers, as the author would agree, I am sure. So at the end we are left just as puzzled about these problems as we might have been before reading the book. Indeed, some readers may have gained the insight needed to be more puzzled than they were at the beginning. As the dust jacket usefully explains: "Davies concludes that in spite of decades of progress in unravelling the mysteries of time, the revolution begun by Einstein remains tantalizingly incomplete". □

Peter Landsberg is at the Faculty of Mathematical Studies, University of Southampton, Southampton SO9 5NH UK.

Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits

Robert McDaniel^{*}, Susanne Ebert-Khosla^{*}, David A. Hopwood[†]
& Chaitan Khosla^{*‡}

^{*} Department of Chemical Engineering, Stanford University, Stanford, California 94305-5025, USA

[†] Department of Genetics, John Innes Centre, Norwich NR4 7UH, UK

Recent advances in understanding of bacterial aromatic polyketide biosynthesis allow the development of a set of design rules for the rational manipulation of chain synthesis, reduction of keto groups and early cyclization steps by genetic engineering. The concept of rational design is illustrated by the preparation of *Streptomyces* strains that produce two new polyketides by expression of combinations of appropriate enzymatic subunits from naturally occurring polyketide synthases. The potential for generating molecular diversity within this class of molecules by genetic engineering is enormous.

POLYKETIDES are a huge family of structurally diverse natural products with a wide range of biological activities. Some notable examples include tetracyclines and erythromycin (antibiotics), FK506 (immunosuppressant), doxorubicin (anticancer), monensin and avermectin (antiparasitics), all from the bacterial group that produces a large proportion of such compounds, the actinomycetes. These molecules are synthesized by multifunctional polyketide synthase enzymes (PKSs), which catalyse repeated condensation cycles between acyl thioesters (usually acetyl, propionyl, malonyl or methylmalonyl). Each condensation cycle results in the formation, on a growing carbon chain, of a β -keto group that may then undergo none, some or all of a series of reductive steps. PKSs incorporate enormous structural diversity into their products in addition to varying this cycle by controlling the overall chain length, choice of primer and extender units and, particularly in the case of aromatic polyketides, regiospecific cyclizations of the nascent polyketide chain^{1,2}.

Given the clinical and commercial success of polyketides, their shared mechanism of biosynthesis, and the high degree of conservation that exists among PKS gene clusters, considerable interest has developed in the possibility of generating novel polyketides through the rational design of PKSs, or through the construction of combinatorial libraries of genetically engineered PKSs. This concept was fostered by the early success in producing the 'hybrid' polyketides mederrhodin A and dihydrogranatirhodin³. These molecules were generated by transferring partial or complete biosynthetic gene clusters between different polyketide-producing strains and screening for new metabolites. Generation of novel molecules in this manner was largely empirical, and involved the activity of enzymes for late 'tailoring' steps in the biosynthetic pathways rather than the early steps catalysed by the PKSs. Later, Donadio *et al.*^{4,5} demonstrated a rational approach for producing novel macrolides by generating analogues of the erythromycin polyketide backbone through the deletion of active sites within the PKS itself, each eliminating a step in a specific reductive cycle.

An approach to the rational design of novel aromatic polyketides cannot, however, be so straightforward. This is due to fundamental differences in the mechanisms by which aromatic and macrolide PKSs control product structure. Macrolide PKSs

contain a unique active site for each enzyme-catalysed reaction in the pathway⁴. Therefore, product structure is determined by the number and types of active sites. Aromatic PKSs, on the other hand, seem to contain only a single set of iteratively used active sites, either distributed over a single protein (in the fungal enzymes)⁶, or, in the bacterial examples, carried on separate proteins⁷⁻⁹. The basic architecture of the gene clusters encoding the sets of subunits of bacterial aromatic PKSs is very similar (Fig. 1), and so prediction of polyketide structure from the gene structure is not possible. Although sequence analysis has been useful in identifying the functions of key aromatic PKS components, it falls short of predicting the specificities of these enzymes and of elucidating the unknown functions of several cyclizing and modifying enzymes that occur in this class of PKSs.

We have reported the construction of a specially designed host-vector system for expression of engineered PKS gene clusters¹⁰. It consists of *Streptomyces coelicolor* strain CH999, from which the *act* aromatic PKS gene cluster (Fig. 1) and other *act* genes have been deleted, together with the expression plasmid pRM5. This system has already been used to generate many novel products of aromatic and macrolide PKSs¹⁰⁻¹⁹. By relating product structure to the subunit composition of the recombinant aromatic PKSs, it has been possible to ascribe the functions and specificities to the various subunits. For example, proteins involved in determining chain length, the degree and regiospecificity of ketoreduction, and the regiospecificity of cyclizations and aromatizations have been identified¹⁰⁻¹⁴. These studies, which began in a largely empirical mode, have now reached the point of revealing strategies for the *rational* design of novel aromatic polyketides, as we show here by the production, 'to order', of the two new polyketides SEK43 and SEK26. Taken together with earlier results, these advances enable us to propose a set of guidelines for engineering novel polyketides in a predictive manner. We also evaluate the potential for generating molecular diversity by *in vivo* combinatorial biosynthesis of aromatic polyketides.

Rationally designed polyketides

All identified gene clusters for actinomycete aromatic polyketides contain a set of three genes encoding a so-called 'minimal PKS'. This consists of a ketosynthase (KS), which also carries a putative acyltransferase (AT) domain, a chain length factor (CLF), and an acyl carrier protein (ACP) (Fig. 1). It is feasible

[‡] To whom correspondence should be addressed.

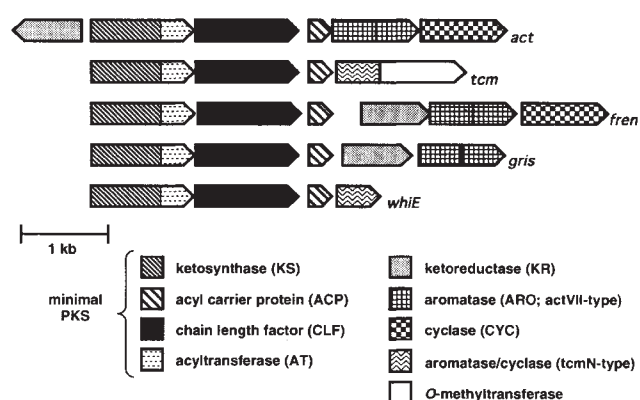


FIG. 1 Genes for related bacterial aromatic polyketide synthases. Relevant portions of the gene clusters responsible for biosynthesis of the aromatic polyketides that give rise to actinorhodin (*act*) in *Streptomyces coelicolor*^{9,30}, tetracenomycin (*tcm*) in *Streptomyces glaucus*^{7,31}, frenolicin (*fren*) in *Streptomyces roseofulvus*²², griseusin in *Streptomyces griseus*²⁰, and the *whiE*-encoded spore pigment in *S. coelicolor*³². Each cluster contains genes encoding a minimal PKS, consisting of a KS/putative AT, a CLF and an ACP, which are collectively responsible for synthesis of the polyketide backbone. The *act*, *fren* and *gris* clusters also contain KR and ARO genes whose products catalyse regiospecific ketoreduction and aromatization of the nascent polyketide chain. The *act*, *fren* and *gris* aromatizing subunits are postulated to have been derived from an internal gene duplication to yield similar N- and C-terminal domains²². The *act* and *fren* clusters also contain CYC genes responsible for formation of the second ring. The *tcm* cluster encodes the bifunctional TcmN protein whose N-terminal domain is involved in aromatization and cyclization reactions. The same is true of the corresponding *whiE* protein.

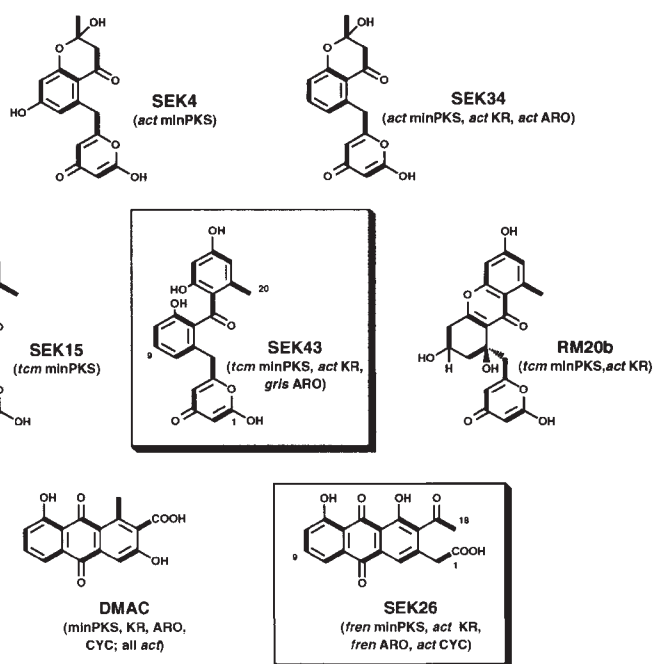
to build a 16-carbon molecule, for example SEK4 (Fig. 2), from the *act* minimal PKS alone. To produce the C-9 reduced analogue of SEK4, SEK34 (Fig. 2), two additional activities are needed: a ketoreductase (KR) and an aromatizing subunit (ARO) (compare the genes present on pSEK24 and pSEK34; Table 1). We reasoned that analogous pairs of molecules could be generated from backbones of different chain length, for example 20 carbons. The challenge was to identify a suitable combination of a minimal PKS, a KR and an ARO.

Earlier studies had established that the *tcm* minimal PKS (on pSEK33; Table 1) is both necessary and sufficient for synthesis of an unreduced 20-carbon backbone, which forms SEK15 (Fig. 2)¹². It was also known that the *act* KR can reduce the C-9 carbonyl on such a backbone to a hydroxyl. Although aromatization of the unreduced ring occurs spontaneously, aromatization of the reduced ring requires an ARO, whose action removes

the hydroxyl^{13,14}. However, previous studies indicated that the *act* ARO cannot aromatize 20-carbon chains^{10,12}. Furthermore, the *tcm* PKS cluster (which lacks a KR gene) does not appear to encode a first ring ARO which would be a suitable candidate. Instead, an ARO gene, homologous to the one in the *act* cluster, was chosen from the gene cluster (Fig. 1) that encodes the PKS for the 20-carbon polyketide griseusin (*gris*)²⁰. The plasmid pSEK43 (Table 1), containing the *tcm* minimal PKS, the *act* KR, and the *gris* ARO, was constructed and introduced into the CH999 host. Analysis of the transformed strain revealed the desired polyketide SEK43, whose structure was determined by NMR, mass spectroscopy and isotopic labelling studies (Fig. 2).

The biosynthesis of SEK43 (Fig. 3) also reaffirms our previous hypothesis that the *act* ARO and its homologues aromatize the first ring¹³. Without a functional ARO, the *tcm* minimal PKS and *act* KR (pSEK23; Table 1) produce RM20b (Fig. 2), which

FIG. 2 Structures of the two novel polyketides SEK43 and SEK26 (boxed) and other relevant polyketides produced by genetic engineering in *S. coelicolor* CH999. The acetate labelling patterns in the polyketide backbones are illustrated with bold bonds. Molecules with 16, 18 and 20 carbons are derived from 8, 9 and 10 acetate units, respectively. Experimental procedures for the sodium [1,2-¹³C₂] acetate incorporation, purification and characterization of these compounds are analogous to those described elsewhere^{11–15}. The ¹H- and ¹³C-NMR spectra of SEK43 are similar to those of its previously characterized analogue SEK15¹⁴. Likewise, spectra for SEK26 were similar to those of DMAC¹¹. Nuclear Overhauser effect measurements were consistent with the proposed structures of both molecules. Coupling constants for the ¹³C spectra refer to carbon–carbon couplings obtained through [1,2-¹³C₂] acetate incorporation. Carbons are numbered sequentially according to the polyketide backbone, starting from the carboxyl end. Symbols used as follows: δ , chemical shift; s, singlet; dd, doublet of doublets; d, doublet; J , ¹H–¹H coupling constant; J_{CC} , ¹³C–¹³C coupling constant. SEK43: ¹H NMR (400-MHz, DMSO-*d*₆) δ (p.p.m.) 12.72 (s, 10H), 11.56 (s, 10H), 10.45 (s, 10H), 9.82 (s, 10H), 7.22 (dd, J =7.6, 8.0 Hz, 1H, C-9), 6.80 (d, J =8.3 Hz, 1H, C-8), 6.77 (d, J =7.9 Hz, 1H, C-10), 6.13 (s, 1H, C-18), 6.09 (s, 1H, C-16), 5.69 (s, 1H, C-4), 5.14 (s, 1H, C-2), 3.58 (s, 2H, C-6), 1.83 (s, 3H, C-20); ¹³C-NMR (100-MHz, DMSO-*d*₆) δ (p.p.m.) (J_{CC} (Hz), carbon) 163.7 (79.1, C-1), 88.6 (79.0, C-2), 170.3 (57.6, C-3), 101.2 (57.7, C-4), 163.8 (51.7, C-5), 36.6 (51.4, C-6), 132.7 (58.8, C-7), 121.1 (58.5, C-8), 130.2 (56.7, C-9), 114.8 (56.7, C-10), 153.9 (69.2, C-11), 130.9 (69.4, C-12), 200.0 (58.3, C-13), 115.7 (59.5, C-14), 165.3 (68.2, C-15), 100.8 (67.8, C-16), 163.3 (61.1, C-17), 111.7 (61.3, C-18), 143.1 (42.0, C-19), 21.6 (42.3, C-20). Field desorption mass spectroscopy gave ion mass-to-charge ratios of 340 and 296 (decarboxylated product). SEK26: ¹H NMR (400-MHz, DMSO-*d*₆) δ (p.p.m.) 12.39 (s, 10H), 11.78 (s, 10H), 7.84 (dd, J =8.1, 7.7 Hz, 1H), 7.74 (d, J =7.5 Hz, 1H), 7.70 (s, 1H), 7.42 (d, J =8.3 Hz, 1H), 3.80



(s, 2H), 2.56 (s, 3H); ¹³C-NMR (100-MHz, DMSO-*d*₆) δ (p.p.m.) (J_{CC} (Hz), carbon) 171.1 (54.5, C-1), 38.4 (54.1, C-2), 141.6 (56.5, C-3), 121.9 (56.9, C-4), 133.1 (54.3, C-5), 181.1 (54.5, C-6), 133.3 (61.7, C-7), 119.6 (61.8, C-8), 137.7 (56.7, C-9), 124.6 (56.9, C-10), 161.3 (63.4, C-11), 116.1 (63.6, C-12), 191.8 (56.3, C-13), 115.4 (56.1, C-14), 158.3 (67.0, C-15), 135.8 (67.4, C-16), 202.6 (42.1, C-17), 31.5 (41.9, C-18). High-resolution fast atom bombardment mass spectroscopy gave an ion mass-to-charge ratio of 369.0974 ($M + H^+$, 369.0974 expected).

TABLE 1 Polyketides produced by recombinant PKSs

Plasmid	Minimal PKS*	KR	ARO	CYC	Product	Reference
pSEK24	<i>act</i>	—	—	—	SEK4	12
pSEK34	<i>act</i>	<i>act</i>	<i>act</i>	—	SEK34	13
pRM5	<i>act</i>	<i>act</i>	<i>act</i>	<i>act</i>	DMAC	10
pSEK33	<i>tcm</i>	—	—	—	SEK15	12
pSEK23	<i>tcm</i>	<i>act</i>	—	—	RM20b	12
pSEK41	<i>act</i>	<i>act</i>	<i>gris</i>	—	SEK34	This work
pSEK42	<i>fren</i>	<i>act</i>	<i>gris</i>	—	SEK34	This work
pSEK43†	<i>tcm</i>	<i>act</i>	<i>gris</i>	—	SEK43	This work
pSEK25	<i>act</i>	<i>act</i>	<i>fren</i>	<i>act</i>	DMAC	This work
pSEK26‡	<i>fren</i>	<i>act</i>	<i>fren</i>	<i>act</i>	DMAC/SEK26	This work
pSEK27	<i>tcm</i>	<i>act</i>	<i>fren</i>	<i>act</i>	RM20b	This work
pRM51	<i>fren</i>	<i>act</i>	<i>gris</i>	<i>act</i>	DMAC/SEK26	This work
pRM52	<i>tcm</i>	<i>act</i>	<i>gris</i>	<i>act</i>	SEK43	This work
pSEK44	<i>act</i>	—	<i>gris</i>	—	SEK4	This work
pSEK45	<i>tcm</i>	—	<i>gris</i>	—	SEK15	This work
pSEK46	<i>act</i>	—	<i>fren</i>	<i>act</i>	SEK4	This work
pSEK47	<i>tcm</i>	—	<i>fren</i>	<i>act</i>	SEK15	This work
pRM53	<i>act</i>	—	<i>gris</i>	<i>act</i>	SEK4	This work
pRM54	<i>tcm</i>	—	<i>gris</i>	<i>act</i>	SEK15	This work

Materials and methods for the manipulation of plasmids and bacterial strains are identical to those described elsewhere^{11–13}. Both the *fren* and *gris* ARO genes were amplified by polymerase chain reaction (PCR) with *Pst*I and *Bsp*HI restriction sites flanking the 5' and 3' ends, respectively. The genes were then subcloned into the *Pst*I/*Bsp*HI sites of pJ5639²¹, displacing the *act*VII gene which encodes the *act* ARO. The *Xba*I-*Eco*RI fragments of the resulting plasmids, containing genes for the *act* ACP, *act* CYC, and *fren* or *gris* ARO genes, were cloned into: pRM34¹¹ to generate pSEK26 and pRM51; pRM37¹⁰ to generate pSEK27 and RM52; pSEK4¹⁰ to generate SEK46 and RM53; and pSEK15¹⁰ to generate pSEK47 and RM54. The subclone containing the *fren* CYC was also introduced into pRM5¹⁰ to generate pSEK25. To construct the plasmids pSEK41–45, a *Kpn*I fragment from pJ5612²⁰ containing the *gris* CYC gene was cloned into the *Pst*I-*Eco*RI sites of pJ5639, displacing the *act* ARO and *act* CYC genes. The *Xba*I-*Eco*RI fragment of the resulting plasmid, containing the *act* ACP and *gris* ARO genes, was cloned into the corresponding sites of pSEK34, pRM5, pRM37, pSEK4 and pSEK15 to yield pSEK41–45, respectively.

* The minimal PKS consists of the ketosynthase/putative acyl transferase and the chain length factor from the gene cluster indicated, together with the *act* acyl carrier protein (Fig. 1).

† Although this recombinant strain produced several minor products (<1 mg l⁻¹), SEK43 was the most abundant product isolated via the work-up procedures used in this study. The yield of SEK43 was ~100 mg l⁻¹. More importantly, neither RM20b nor SEK15 were detected by HPLC in CH999/pSEK43. Likewise, SEK43 was not detected by HPLC in strains expressing the *tcm* minimal PKS and the *act* KR but lacking the *gris* ARO (for example, CH999/pSEK27), neither was it detected in strains expressing the *tcm* minimal PKS and the *gris* ARO but lacking the *act* KR (for example, CH999/pSEK45). Taken together, the above results imply that the rational design and engineered biosynthesis of SEK43 is a direct result of the combination of all three PKS components (that is, the *tcm* minimal PKS, the *act* KR and the *gris* ARO).

‡ Although this recombinant strain produced several minor products (<1 mg l⁻¹), DMAC and SEK26 were the most abundant octaketides and nonaketides, respectively, isolated via the work-up procedures used in this study. The yield of DMAC was ~25 mg l⁻¹, and that of SEK26 was ~50 mg l⁻¹. More importantly, RM18¹¹ was not detected by HPLC in CH999/pSEK26. RM18 is the major nonaketide product of a strain containing the *fren* minimal PKS, *act* KR, *act* ARO and *act* CYC¹¹, as well as a strain containing the *fren* minimal PKS and *act* KR only¹². Likewise, SEK26 was present only in trace amounts (<1 mg l⁻¹) in the control strains which lack the *fren* ARO and/or the *act* CYC. Taken together, the above results imply that the rational design and engineered biosynthesis of SEK26 is a direct result of the combination of all four PKS components (that is, the *fren* minimal PKS, the *act* KR, the *fren* ARO and the *act* CYC).

contains a non-aromatized first ring. Replacement of the *tcm* minimal PKS in pSEK43 with either the *act* or *fren* minimal PKSs (pSEK41 and pSEK42; Table 1) resulted in production of the 16-carbon aromatized compound SEK34, demonstrating that the *gris* ARO can also recognize shorter carbon chains. It was unexpected, however, that a corresponding 18-carbon polyketide was not detected in the construct containing the *fren* minimal PKS, which has been shown to synthesize both 18- and 16-carbon chains¹¹. This is probably due to decomposition of the molecule, as CH999/pSEK42 produced small quantities of an uncharacterized molecule not present in CH999/pSEK34. More significantly, evidence for an aromatized 18-carbon intermediate is described below.

A second test of the concept of rational design arose from the previous isolation of the 16-carbon polyketide DMAC (Fig. 2). The PKS subunits required for DMAC biosynthesis are a '16-carbon' minimal PKS, a KR, and suitable ARO and CYC components (pRM5; Table 1). Here, the CYC catalyses cyclization of the *second* ring between carbons 4 and 15, leading eventually to the formation of an anthraquinone¹³. These observations suggested that an analogous anthraquinone, with 18 carbons, could be generated. To achieve this, the plasmid pSEK26 (Table 1) containing the *fren* minimal PKS with the *act* KR, the *fren* ARO, and the *act* CYC was constructed. The *fren* minimal PKS and

act KR were selected for their ability to produce an 18-carbon, C-9 reduced backbone^{11,12}. The *fren* ARO was chosen because the *act* ARO cannot aromatize 18-carbon chains^{11,12}. Introduction of the plasmid into CH999 resulted in the production of both DMAC and SEK26. The latter is indeed a novel 18-carbon anthraquinone whose structure was confirmed by NMR, mass spectroscopy and isotopic labelling studies (Fig. 2). Formation of SEK26 (Fig. 3) occurs through a second ring cyclization at C5/C14 presumably catalysed by the *act* CYC. The production also of DMAC is consistent with the relaxed chain-length specificity of the *fren* minimal PKS¹¹.

To evaluate further the specificity of ARO and CYC subunits towards carbon chains of varying length, several other PKS combinations were constructed (Table 1). For example, pSEK25 and pSEK26 demonstrate that the *fren* ARO can aromatize both 16- and 18-carbon chains. However, the *fren* ARO cannot handle 20-carbon chains; instead the combination of the *fren* ARO with the *tcm* minimal PKS, *act* KR, and *act* CYC (pSEK27) resulted in biosynthesis of RM20b, the non-aromatized 20-carbon polyketide (Fig. 2). As expected, replacing the *fren* ARO with the *gris* ARO (pRM51) in pSEK26 yielded DMAC and SEK26. But attempts to generate a 20-carbon polyketide with a C-5/C-14 second ring cyclization were unsuccessful; replacing the *fren* minimal PKS in pRM51 with the *tcm* minimal PKS (pRM52)

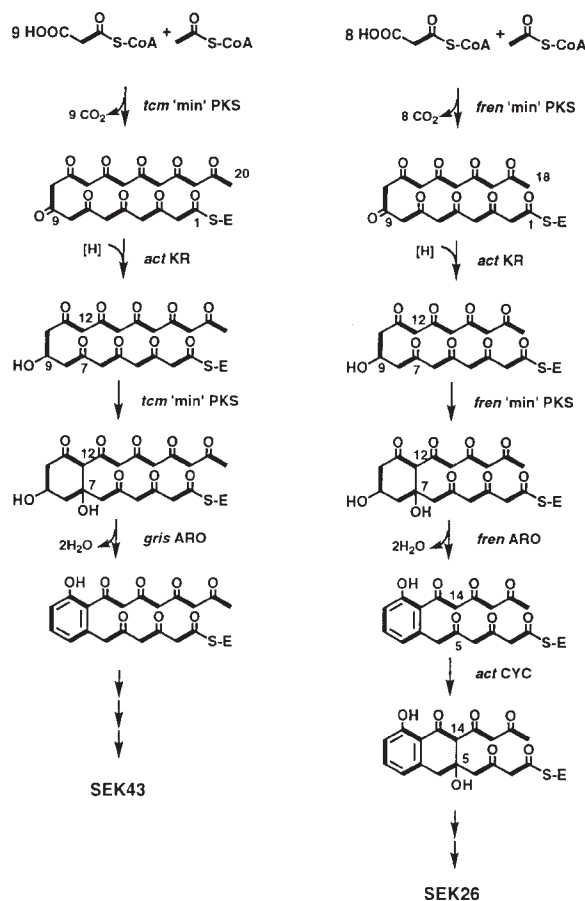


FIG. 3 Proposed biosynthetic pathways for the rationally designed polyketides, SEK43 and SEK26. For details, see text.

resulted in production of SEK43, indicating that the *act* CYC cannot cyclize 20-carbon chains. Finally, the plasmids pSEK44–47, pRM51 and pRM52 (Table 1), all lacking KRs, failed to cause production of polyketides different from those produced by the minimal PKS alone¹², despite the presence of ARO and CYC components. This is consistent with previous observations that ARO and CYC subunits do not alter the biosynthetic pathways of unreduced polyketides^{12,15}.

Design rules for engineered biosynthesis

As described above, analysis of the products of many combinations of bacterial aromatic PKS components has led to significant advances in our understanding of aromatic PKS function and specificity, including the constraints that limit the activity of specific subunits to a potential range of substrates. Based on these results, we present a series of conclusions that may be considered as design rules for manipulating the early steps of aromatic polyketide biosynthesis.

(1) Chain length. Polyketide carbon-chain length is dictated by the minimal PKS (Fig. 4)^{10–12}. Within the minimal PKS, the acyl carrier protein can be interchanged without affecting specificity^{10–12,21}, whereas the chain length factor is crucial. Although some ketosynthase/chain length factor combinations are functional, others are not^{10,11}; therefore, biosynthesis of a polyketide chain of specified length can be ensured with a minimal PKS in which both the ketosynthase and chain length factor originate from the same PKS gene cluster. So far, chain lengths of 16 (octaketide), 18 (nonaketide), 20 (decaaketide) and 24 carbons (dodecaaketide) can be generated with minimal PKSs from the *act*, *fren*, *tcm* and *whiE* PKS clusters, respectively (refs 10–12 and T.-W. Yu, R.M., C.K. and D.A.H., manuscript in

preparation). At least in the presence of a KR, the *whiE* minimal PKS can also generate 22-carbon backbones, suggesting a degree of relaxed chain length control as found for the *fren* PKS (T.-W. Yu, R.M., C.K. and D.A.H., manuscript in preparation).

(2) Ketoreduction. Ketoreduction requires a ketoreductase (Fig. 4). The *act* KR (the only one studied so far) can catalyse reduction of the C-9 carbonyl (counting from the carboxyl end) of any length nascent polyketide backbone studied so far¹⁴. Furthermore, the *act* KR is compatible with all the minimal PKSs mentioned above¹². Homologous ketoreductases have been identified in other PKS clusters^{8,20,22}, and these enzymes can also probably catalyse ketoreduction at C-9, as all the corresponding natural products undergo this modification. In unusual circumstances, C-7 ketoreductions have also been observed with the *act* KR¹¹.

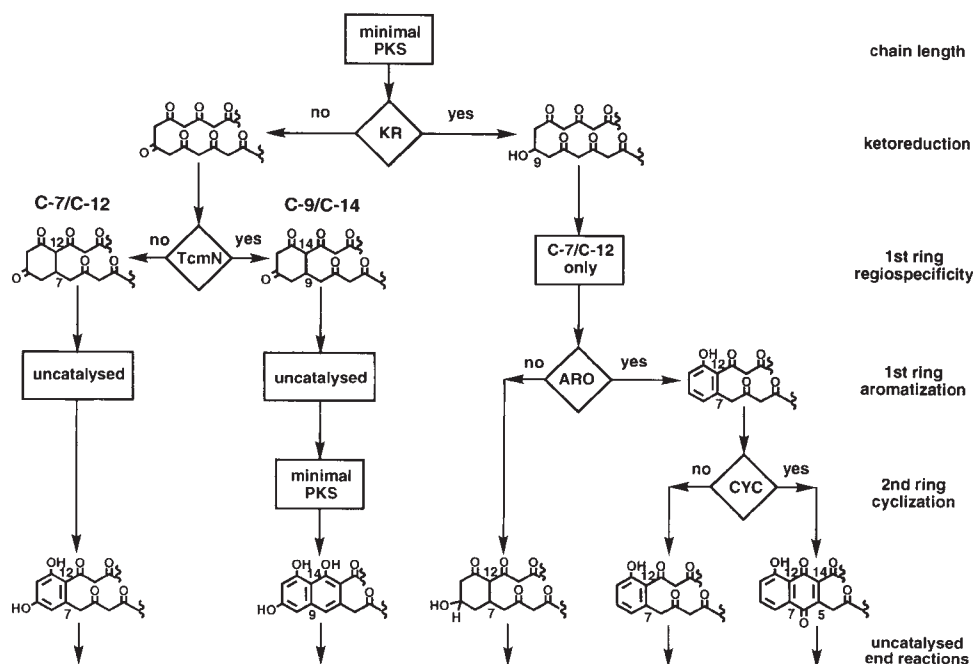
(3) Cyclization of the first ring. Although the minimal PKS alone can control formation of the first ring, the regiospecific course of this reaction may be influenced by other PKS proteins. For example, most minimal PKSs studied so far produce polyketides with C-7/C-12 cyclizations when present alone (Fig. 4). In contrast, the *tcm* minimal PKS alone generates both C-7/C-12 and C-9/C-14 cyclized products¹². The presence of a ketoreductase with any minimal PKS restricts the nascent polyketide chain to cyclize exclusively with respect to the position of ketoreduction: C-7/C-12 cyclization for C-9 ketoreduction, and C-5/C-10 cyclization for C-7 ketoreduction (Fig. 4)^{11–13}. Likewise, use of the TcmN enzyme alters the regiospecificity to C-9/C-14 cyclizations for unreduced polyketides of different lengths, but has no effect on reduced molecules (Fig. 4)²³.

(4) First ring aromatization. The first ring in unreduced polyketides aromatizes non-catalytically. In contrast, an aromatizing subunit is required for reduced polyketides (Fig. 4)¹³. There seems to be a hierarchy in the chain-length specificity of these subunits from different PKS clusters. For example, the *act* ARO will recognize only 16-carbon chains¹², the *fren* ARO recognizes both 16- and 18-carbon chains, whereas the *gris* ARO recognizes chains of 16, 18 and 20 carbons.

(5) Second ring cyclization. C-5/C-14 cyclization of the second ring of reduced polyketides may be achieved with an appropriate cyclase (Fig. 4)¹³. Although the *act* CYC can cyclize octa- and nonaketides, it does not recognize longer chains. No equivalent C-5/C-14 cyclase with specificity for decaaketides or longer chains has been identified, although the structures of natural products such as griseusin imply their existence. In the case of sufficiently long unreduced chains with a C-9/C-14 first ring, formation of a C-7/C-16 second ring is catalysed by the minimal PKS (Fig. 4)^{12,23}.

(6) Additional cyclizations. The KS, CLF, ACP, KR, ARO and CYC subunits of the PKS together catalyse the formation of an intermediate with a defined chain length, reduction pattern, and first two cyclizations. Although the biosynthesis of naturally occurring polyketides typically requires the activity of downstream cyclases and other modifying enzymes to generate the characteristic biologically active product, subsequent reactions in the biosynthesis of engineered polyketides described here and in our earlier work occur in the absence of specific enzymes and are determined by the different physical and chemical properties of the individual molecules. Presumably reflecting such chemical possibilities and constraints, consistent patterns have been observed, leading to some degree of predictability. Two common moieties formed by the uncyclized methyl terminus of polyketide chains are hemiketals and benzene rings. Formation of a hemiketal occurs in the presence of an appropriately positioned enol and can be followed by a dehydration because both the hydrated and dehydrated forms are often isolated (Fig. 5a)^{10,13,14,23}; benzene-ring formation occurs with longer unprocessed methyl ends (Fig. 5b)¹⁴. The most frequently observed moiety at the carboxyl terminus of the chain is a γ -pyrone ring formed by three ketide units (Fig. 5c)^{13,15,17,24}; if a free carboxylic acid remains, decarboxylation typically occurs if a β -carbonyl exists

FIG. 4 Scheme for rationally engineered biosynthesis of polyketides. Current degrees of freedom for the manipulation of polyketide structure include chain length, ketoreduction, regiospecificity of the first ring, aromatization of the first ring and cyclization of the second ring. See text for details and Fig. 5 for uncatalysed end reactions.



(Fig. 5d)^{10,11,19}. Many aldol condensations can be predicted as well, bearing in mind that the methyl and carboxyl ends tend preferentially to cyclize independently but will co-cyclize if no alternative exists (Fig. 5e)¹². These non-enzymatic cyclization patterns observed *in vivo* are also consistent with earlier biomimetic studies²⁵.

Combinatorial biosynthetic potential

Taken together with the structures of other naturally occurring bacterial aromatic polyketides, the design rules presented above can be extrapolated to estimate the extent of molecular diversity that might be generated by *in vivo* combinatorial biosynthesis. For this calculation, we consider separately the cases of reduced and unreduced polyketides. For reduced polyketides, the identified degrees of freedom include chain length, aromatization of the first ring, and cyclization of the second ring. For unreduced polyketides, these include chain length and regiospecificity of the first ring cyclization. The number of accessible structures is the product of the number of ways in which each degree of freedom can be varied. Chains of five different lengths have so far been manipulated. From the structure and deduced biosynthetic pathways of the dynemicin anthraquinone²⁶, simaomicin²⁷ and benastatin²⁸, we anticipate the isolation of minimal PKSs that generate 14, 26- and possibly 28-carbon backbones, respectively, bringing the potential number to eight. Use of the genes for already isolated minimal PKSs, such as the *actI* genes, can be a powerful route to the cloning of such minimal PKSs^{8,20,22,29}. Reduced chains can either be aromatized or not; likewise, a second ring cyclase is optional in cases where the first ring is aromatized (Fig. 4). The regiospecificity of the first cyclization of an unreduced chain can be varied, depending on the presence of an enzyme like TcmN. Thus, the numbers are $8 \times 3 = 24$ for reduced chains and $8 \times 2 = 16$ for unreduced chains. Although this is a very rudimentary estimate, in practice the numbers are likely to be significantly higher because of the presence of several additional minor products (which can be of the order of 5–10 per major product) that result from as-yet uncharacterized pathways.

The number of potential novel polyketides should increase geometrically as new degrees of freedom are exploited and/or protein engineering strategies are brought to bear on the task of creating enzyme subunits with specificities not observed in

nature. For example, non-acetate starter units can be incorporated into polyketide backbones (for example, propionate in daunorubicin and malonamide in oxytetracycline), although the basis for this is currently obscure. Then again, enzymes that catalyse downstream cyclizations and late-step modifications, such as group transfer reactions and oxidoreductions commonly seen in naturally occurring polyketides, can be studied along the lines presented here and elsewhere^{10–17}. It is therefore possible that at least some of these degrees of freedom can be combinatorially exploited in the laboratory to generate libraries of 'unnatural' natural products with structural diversity that is comparable to that observed in nature. □

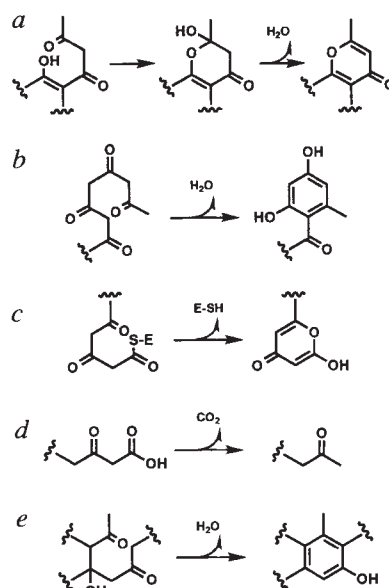


FIG. 5 Common moieties observed in engineered polyketides. These structures are formed by non-enzymatic reactions involving the uncyclized portions of the carbon chain. Hemiketals (a) and benzene rings (b) are formed at the methyl ends, whereas γ -pyrone rings (c) and decarboxylations (d) occur at the carboxyl ends. The two chain ends can also co-cyclize via aldol condensations (e). See text for details.

Received 6 March; accepted 9 May 1995.

1. O'Hagan, D. *The Polyketide Metabolites* (Ellis Horwood, Chichester, UK, 1991).
2. Katz, L. & Donadio, S. A. *Rev. Microbiol.* **47**, 875–912 (1993).
3. Hopwood, D. A. et al. *Nature* **314**, 642–644 (1985).
4. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. *Science* **252**, 675–679 (1991).
5. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. *Gene* **115**, 97–103 (1992).
6. Beck, J., Ripka, S., Siegner, A., Schiltz, E. & Schweizer, E. *Eur. J. Biochem.* **192**, 487–498 (1990).
7. Bibb, M. J., Biró, S., Motamedi, H., Collins, J. F. & Hutchinson, C. R. *EMBO J.* **8**, 2727–2736 (1989).
8. Sherman, D. H. et al. *EMBO J.* **8**, 2717–2725 (1989).
9. Fernández-Moreno, M. A., Martínez, E., Boto, L., Hopwood, D. A. & Malpartida, F. *J. Biol. Chem.* **267**, 19278–19290 (1992).
10. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *Science* **262**, 1546–1550 (1993).
11. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *J. Am. chem. Soc.* **115**, 11671–11675 (1993).
12. McDaniel, R., Ebert-Khosla, S., Fu, H., Hopwood, D. A. & Khosla, C. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11542–11546 (1994).
13. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *J. Am. chem. Soc.* **116**, 10855–10859 (1994).
14. Fu, H., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *J. Am. chem. Soc.* **116**, 4166–4170 (1994).
15. Fu, H., McDaniel, R., Hopwood, D. A. & Khosla, C. *Biochemistry* **33**, 9321–9326 (1994).
16. Fu, H., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. *J. Am. chem. Soc.* **116**, 6443–6444 (1994).
17. Fu, H., Hopwood, D. A. & Khosla, C. *Chem. Biol.* **1**, 205–210 (1994).
18. Kao, C. M., Katz, L. & Khosla, C. *Science* **265**, 509–512 (1994).
19. Kao, C. M., Luo, G., Katz, L., Cane, D. E. & Khosla, C. *J. Am. chem. Soc.* **116**, 11612–11613 (1994).
20. Yu, T.-W., Bibb, M. J., Revill, P. W. & Hopwood, D. A. *J. Bacteriol.* **176**, 2627–2634 (1994).
21. Khosla, C. et al. *J. Bacteriol.* **175**, 2197–2204 (1993).
22. Bibb, M. J., Sherman, D. H., Omura, S. & Hopwood, D. A. *Gene* **142**, 31–39 (1994).
23. McDaniel, R., Hutchinson, C. R. & Khosla, C. *J. Am. chem. Soc.* (in the press).
24. Zhang, H.-L. et al. *J. org. Chem.* **55**, 1682–1684 (1990).
25. Griffin, D. A., Leeper, F. J. & Staunton, J. J. *chem. Soc., Perkin Trans. 1*, 1035–1042 (1984).
26. Tokiwa, Y. et al. *J. Am. chem. Soc.* **114**, 4107–4110 (1992).
27. Carter, G. T., Goodman, J. J., Torrey, M. J., Borders, D. B. & Gould, S. J. *J. org. Chem.* **54**, 4321–4323 (1989).
28. Aoyama, T., Naganawa, H., Muraoka, Y., Aoyagi, T. & Takeuchi, T. *J. Antibiot.* **45**, 1767–1772 (1992).
29. Malpartida, F. et al. *Nature* **325**, 818–821 (1987).
30. Hallam, S. E., Malpartida, F. & Hopwood, D. A. *Gene* **74**, 305–320 (1988).
31. Summers, R. G., Wendt-Pienkowski, E., Motamedi, H. & Hutchinson, C. R. *J. Bacteriol.* **174**, 1810–1820 (1992).
32. Davis, N. K. & Chater, K. F. *Molec. Microbiol.* **4**, 1679–1691 (1990).

ACKNOWLEDGEMENTS. This work was supported by a grant from the US NSF, an NSF Young Investigator Award, and a David and Lucile Packard Fellowship for Science and Engineering to C.K. D.A.H. acknowledges financial support from the Biotechnology and Biological Sciences Research Council and the John Innes Foundation.

The 2.2 Å crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue

Nicolas Nassar, Gudrun Horn, Christian Herrmann, Anna Scherer*, Frank McCormick† & Alfred Wittinghofer

Max-Planck-Institut für molekulare Physiologie, Abteilung Strukturelle Biologie, Postfach 102664, 44026 Dortmund, Germany

* Max-Planck-Institut für medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, 69119 Heidelberg, Germany

† Onyx Pharmaceuticals, Richmond, California 94806–5206, USA

The X-ray crystal structure of the complex between the Ras-related protein Rap1A in the GTP-analogue (GppNHp) form and the Ras-binding domain (RBD) of the Ras effector molecule c-Raf1, a Ser/Thr-specific protein kinase, has been solved to a resolution of 2.2 Å. It shows that RBD has the ubiquitin superfold and that the structure of Rap1A is very similar to that of Ras. The interaction between the two proteins is mediated by an apparent central antiparallel β -sheet formed by strands B1–B2 from RBD and strands β 2– β 3 from Rap1A. Complex formation is mediated by main-chain and side-chain interactions of the so-called effector residues in the switch I region of Rap1A.

THE Ser/Thr-specific protein kinase c-Raf1 was originally identified as the oncogenic component of the murine sarcoma virus 3611. The complementary DNA of the human proto-oncogene revealed that the protein contains 648 amino-acid residues which code for an N-terminal regulatory region and a C-terminal kinase domain, while the Raf oncogene is missing part of the regulatory region. Two additional members of the Raf kinase family (called A and B) have been identified and genes for this class of protein kinases have also been found in invertebrates, as *polehole* or *D-raf* in *Drosophila* and *lin-45* in *Caenorhabditis elegans*. Earlier experiments have shown that c-Raf1 (referred to as Raf) is part of a Ras-dependent signalling pathway from receptors to the nucleus, and that Raf lies downstream of Ras in this pathway (reviewed in refs 1, 2).

Raf is one of several downstream targets of Ras because part of the regulatory region of Raf binds to Ras in a GTP-dependent

manner and mutations in the effector region (residues 32–40) of Ras that are deficient in Ras signalling do not bind Raf^{3–8}. The function of this binding in the activation of Raf kinase is to bring it to the plasma membrane^{9,10}, where it can be activated by an unknown mechanism. Several clones have been isolated using a yeast double-hybrid screen that code for overlapping Ras-binding fragments of Raf⁶. The only sequence common to these clones is a polypeptide of 81 amino acids comprising residues 51–131, the Ras-binding domain or RBD, which has been expressed and shown to comprise an independent domain sufficient for GTP-dependent binding of Ras^{11–13}.

Rap1A is a small GTP-binding protein with an effector region identical to Ras and an overall sequence identity of 50%. It has been cloned on the basis of its homology to Ras^{14,15} and binds the same candidate effectors as Ras itself such as GAP^{16,17}, Raf⁷, phosphatidylinositol-3-OH kinase PI(3)K; J. Downward,

TABLE 1 Experimental data on crystal structure determination and refinement

Resolution limit (Å)	Native data refinement									Cumulative
	8.0	4.21	3.42	3.02	2.75	2.56	2.42	2.30	2.20	
R_{sym} (%)	2.6	3.0	3.9	5.5	7.9	9.8	12.2	15.8		3.8
Completeness (%)	95.8	96.7	97.8	98.3	98.7	99.0	98.9	98.8		98.0
Free R -factor (all data, %) [†]	27.3	26.3	27.6	30.4	31.2	36.2	33.9	36.5		29.3
R_{work} (all data, %)	17.7	16.5	20.6	24.7	29.0	31.1	31.7	36.1		22.1
R -factor (all data, %)	18.6	17.5	21.3	25.2	29.2	31.6	31.9	36.1		22.8
Results of the rotation and translation searches [‡]										
Rotation function	$\theta_1 = 147.77^\circ$	$\theta_2 = 20.24^\circ$	$\theta_3 = 140.94^\circ$	Height/r.m.s. = 6.2 σ			Highest false peak = 4.7 σ			
Translation functions	$x = 0.205$	$y = 0.302$	$z = 0.246$	Height/r.m.s. = 8.3 σ			Highest false peak = 6.4 σ			

Truncated Rap1A protein lacking residues 168–186 (called Rap), was obtained as a recombinant protein from *E. coli* by coexpression with a molecular chaperone. Expression and purification were basically as described for Ras (G. H. and V. Pizon, manuscript in preparation). Protein-bound nucleotides, a mixture of GDP and GTP, were exchanged with the non-hydrolysable GTP-analogue GppNHP as described earlier²⁴ for Ras. Bound nucleotide was quantified by HPLC. As with Ras, the truncated protein was fully able to bind and hydrolyse GTP and to bind Raf. The Ras-binding domain of human c-Raf1 (Ser 51–Leu 131) was expressed and purified as described for Ras²⁴. RBD and Rap were concentrated to 20 mg ml⁻¹ each before mixing in stoichiometric amounts. To remove excess protein, the complex was applied to a Superdex-75 column equilibrated with buffer A. Fractions containing Rap–RBD were pooled and concentrated (Amicon filtration, 10K cutoff). Diffraction-quality crystals were grown at room temperature from hanging drops containing 15–20 mg ml⁻¹ complex, 10% monomethylether PEG 5000 (Fluka), 50 mM ammonium sulphate, 5 mM CaCl₂, 2.5 mM MgCl₂ and 0.5 mM Dithioerythritol (DTE), 32.2 mM Tris-HCl, pH 7.6, over reservoirs of 20% monomethylether PEG5000, 0.1 M ammonium sulphate, 10 mM CaCl₂, 5 mM MgCl₂ and 1 mM DTE, 64.4 mM Tris-HCl, pH 7.6. Crystals appeared overnight and grew typically to 0.4 × 0.3 × 0.3 mm³. Data were collected at 4 °C from a single crystal on the wiggler beam line BWB7 (DESY, EMBL) at $\lambda = 0.87$ Å on an imaging plate area detector (1 deg min⁻¹ per frame). 112° were collected and processed (70,621 observations) using XDS (ref. 46) and reduced to 16,375 unique reflections. The data are 98% complete to 2.2 Å (98.9% for the 2.5–2.2 Å shell); when data with $I/\sigma(I) < 3.0$ are discarded, the overall completeness is 84.5% (74% for the 2.5–2.2 Å shell). These crystals belong to space group $P2_12_12_1$, as determined by XDS, with unit-cell dimensions of ($a = 44.2$ Å, $b = 71.9$ Å, $c = 100.2$ Å) and one complex molecule per asymmetric unit. A Wilson plot for data in the range 8.0–2.2 Å showed an average temperature factor of 33.1 Å². The structure of the complex was solved by the molecular replacement method using truncated Ras (residues 1–166) (ref. 24) coordinates as a search model (including all side chains). Data in the resolution range 8.0 to 3.4 Å were used that had been collected with a Siemens/Nicolet area detector using focused X-rays from a GX-18-rotating Cu anode generator (34,628 observed and 7,178 unique reflections to 2.8 Å; $R_{\text{sym}} = 6.9\%$). Search for the rotation–translation function as well as the final rigid-body refinement were done with AMoRe⁴⁷, as implemented in the CCP4 package. The highest peak gave a correlation coefficient of 0.325 and an R -factor of 0.502 (the highest false peak gave a correlation coefficient of 0.247 and an R -factor of 0.526); after rigid-body refinement these values reached 0.335 and 0.499. Model building was done using program O (ref. 25) and refined using simulated annealing as implemented in XPLOR 3.1 (ref. 26) (no cutoff) starting at 4,000 K (0.2-fs step). For model building, two maps to 2.8 Å were calculated using COMBINE (W. Kabsch) and superimposed. The first map was calculated with Fourier coefficients $(2m|F_o| - D|F_{pc}|)$, phases α_c , where m and D are factors to suppress model bias resulting from phasing by partial structures with errors²⁷. The second map was calculated with Fourier coefficients $(m|F_o| - D|F_{pc}|)$, phases α_{pc} . The common positive peaks for the two maps were assigned to the RBD domain. The RBD model was built a few residues at a time and the partial model refined. The correctness of the model building was followed by the evolution of the free R -factor: when in a cycle the free R -factor decreased, a new part was added to the RBD; when it increased, the part built in the previous cycle was graphically reinspected and corrected. One positive peak present between two symmetry-related molecules was refined as a calcium ion. The final model showed this Ca²⁺ ion coordinated to three water molecules (at 2.3, 2.4 and 2.6 Å), and the main-chain carbonyl of the conserved Gly 132 (2.3 Å), the side chain of the conserved Glu 125 (Oε2, 2.7 Å) from RBD, and the side chain of Asp 80 (Oδ1 at 2.5 Å, Oδ2 at 2.7 Å) from a symmetry-related RBD. Resolution was then increased in 0.1 Å steps to 2.2 Å. Appropriate peaks of positive density were assigned to water molecules. When water molecules were incorporated, the model was refined either with the conjugate gradient method or with simulated annealing to 2,000 K. Residues Met 76, Cys 81, Met 83 and Cys 96 from RBD have two different conformations in the crystal. The final model consisted of residues 1–167 of Rap, residues 55–131 of RBD, one GppNHP molecule, one Mg²⁺ ion, one Ca²⁺ ion and 87 water molecules. For the data in the shell 8.0–2.2 Å, the final free R -factor was 29.3%, R_{work} was 22.1%, and the R -factor was 22.8%. The average temperature factor of the final model is 34 Å². The density for residues Asp 47 to Glu 49 and Asp 105 from Rap, and for side chains of His 105 and Lys 106 from RBD, is weak; these residues show also a high-temperature factor suggestive of disorder in the crystal. Details to higher resolution will be presented elsewhere.

* $R_{\text{sym}} = \sum_n \langle |k_j I(h) - I(h)| \rangle / \sum_n I(h)$, where $\langle |k_j I(h) - I(h)| \rangle$ is the average absolute deviation of scaled reflections $k_j I(h)$ from the average $I(h)$ of its symmetry and Friedel-related equivalents j .

† $R_{\text{free}} = \sum_n (|F_o(h) - F_c(h)| / \sum_n F_o(h))$, calculated for all data, where $F_o(h)$ is the observed and $F_c(h)$ the calculated structure factor amplitudes for reflection h . Free R -factor is the value of this parameter for the omitted 10% of data during refinement, and R_{work} is its value for the 90% set used for refinement.

‡ Rotation (eulerian angles) and translation parameters (fractional) are as defined in AMoRe.

unpublished) and Ral-GEF/RGL^{18–20}. Rap1A antagonizes Ras function in the cell: indeed, it was cloned as a protein capable of reverting transformation induced by the K-Ras oncogene²¹. Further evidence for a possible role of Rap1A in Ras signalling comes from the *roughened* mutation in *Drosophila*, which is a mutation of Rap (phenylalanine to leucine at position 157 in *Drosophila*, 158 in mammalian Rap1A) that leads to loss of signalling through Ras in the *sevenless* pathway²², and from reports that Rap1A antagonizes the Ras-dependent activation of Erk1 (ref. 23), the Ras-induced germinal vesicle breakdown and its ability to inhibit coupling to muscarinic Ca²⁺ channels¹⁵. It is thought that these biological properties of Rap1A are a result of competition for the same effectors, and that Rap itself does not activate effectors such as Raf because it is not localized in the plasma membrane. Here we report the crystal structure of the complex between the human RBD from c-Raf1 and human Rap1A in the GppNHP form. The amino-acid residues of Rap1A involved in the complex interaction suggest that the

interaction between Rap and Raf should be similar to that of Ras and Raf.

Structure determination and refinement

Expression, purification and crystallization of the complex between truncated human Rap1A protein lacking residues 168–186 (referred to as Rap) and the Ras-binding domain of human c-Raf1, residues 51–131, are described in Table 1. The structure was solved by molecular replacement using truncated Ras²⁴ as a search model (Table 1).

Ras was mutated to fit the Rap sequence with program O (ref. 25) and the initial model was refined with XPLOR 3.1 (ref. 26). Electron density maps were then calculated using the phases obtained from the Rap model and the observed amplitudes to 2.8 Å using the method developed by Read²⁷ to remove model bias (Table 1). The RBD domain was built a few residues at a time and the resulting model refined. By repeating this procedure, we were able to recognize residues Ser 55 to Leu 131 in

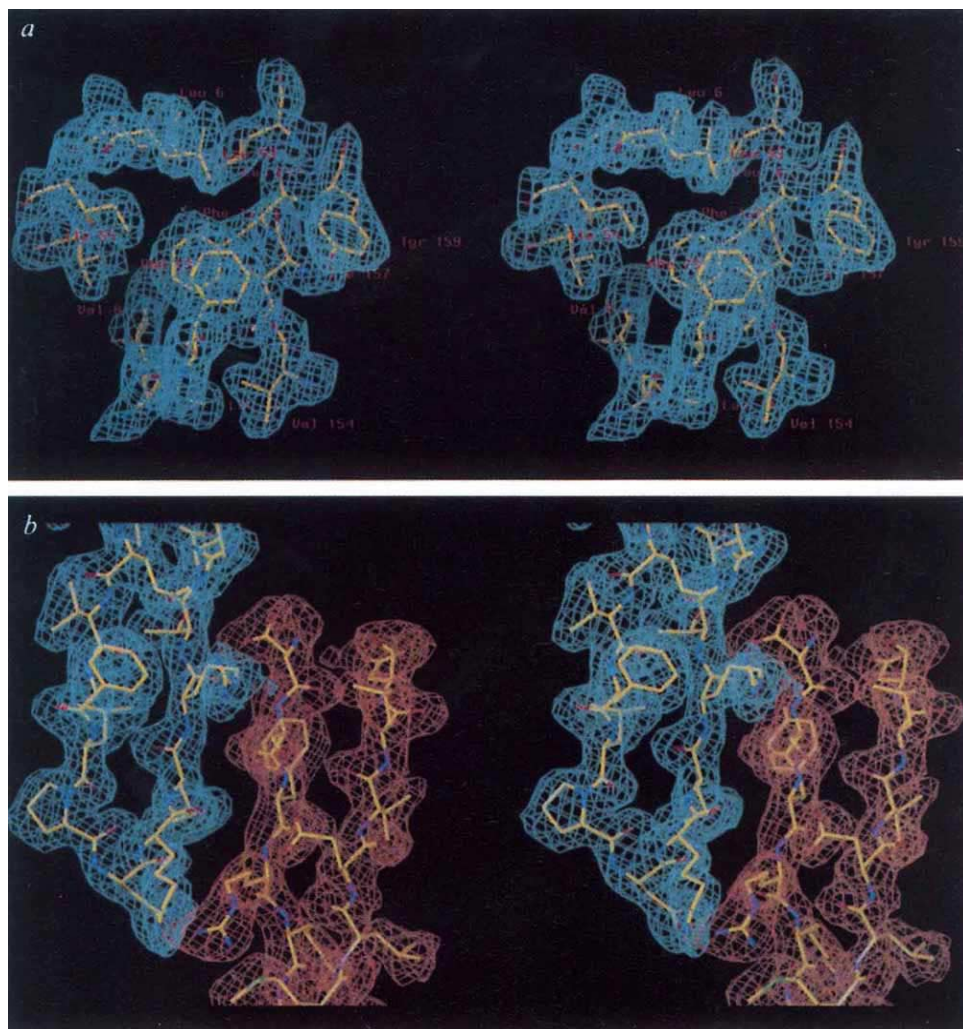


FIG. 1 Final ($2F_o - F_c$) electron density map contoured at 1σ . **a**, Region in Rap surrounding Phe 158, which forms a large hydrophobic core with Leu 6, Val 8, Leu 19, Phe 23, Leu 53, Ile 55, Val 81, Val 154, Ile 157 and Leu 161. Phe 158 corresponds to Phe 157 of *Drosophila* Rap, whose mutation to Leu (Roughened) produces a defect in the Sevenless signalling pathway, which acts via Ras²². **b**, Interacting β -strands β_2 , β_3 of Rap (in orange) and B1, B2 or RBD (in blue).

RBD; residues 51 to 54 are disordered. We believe our model to be correct according to the following criteria: (1) the R -factor of the final model is 22.8% for all data between 8.0 and 2.2 Å; (2) the free R -factor²⁸ for a randomly omitted subset (10%) of data that was at no stage used in the refinement (no cut-off) is 29.3%; (3) the average deviation from standard values is 0.013 Å in bond distance and 2.06° in bond angle; (4) the (Φ - Ψ) plot showed all non-glycine residues within the allowed region apart from Asp 105 of Rap in an disallowed region and His 105 of RBD in a partially allowed region, for neither of which is there any clear electron density; (5) the average real-space correlation coefficient is 0.843, as calculated by O (ref. 25); (6) a $\sigma(A)$ plot gives an error of $<|\Delta r| > = 0.39$ Å in coordinates²⁷; (7) a Luzzati plot gives an upper-limit error on coordinates of 0.40 Å. Two representative segments of the final electron density map (Fig. 1) show the hydrophobic pocket around Phe 158, the site corresponding to the *roughened* mutation in *Drosophila*, which is responsible for interruption of the *sevenless* pathway, and the contacting β -strands of the two proteins. An overall view of the complex is shown in Fig. 2.

Rap

On comparing the structure of RBD-bound Rap in the triphosphate form with uncomplexed Ras in the triphosphate

form^{24,29}, we find these structures to be remarkably similar, with an r.m.s. deviation of 0.88 Å compared to Ras-GppNHp, based on an alpha-carbon ($C\alpha$) superimposition using O (ref. 25). Figure 3 shows an alignment of the primary and secondary structures of Ras and Rap. Whereas the central β -sheet of the G domain is totally conserved, there are slight variations in the helices: $\alpha 3$ is distorted in Ras but not in Rap, there is an extension of $\alpha 4$ at the N terminus (or an additional small helix) which could be caused by one amino-acid insertion (Glu 121), and $\alpha 5$ is longer in Ras. The largest deviations of $C\alpha$ atoms are found for residues Tyr 32 (4.5 Å), Thr 35 and Ile 36 (3 Å) and residues 61–65 from helix $\alpha 2$ (average deviation of 5.7 Å). Whether or not these differences are due to complex formation with RBD is not clear because the three-dimensional structure of Rap by itself is not available. This would be unlikely in the case of $\alpha 2$ as it is not located in the binding interface. The structures of two other GTP-binding proteins from different subfamilies, Arf³⁰ and Ran³¹, show much larger deviations from the structure of the Ras G domain. The close similarity of the structures of Ras and Rap also explains why most of the Ras/Rap chimaeras that have been constructed are biologically active^{32,33}.

Interactions between GppNHp, Mg^{2+} and the protein are also highly conserved between Ras²⁴ and Rap, with a few important differences. In Rap, there are much tighter interactions between

the ribose and the protein which are due to three strong hydrogen bonds between the hydroxyl groups of the ribose and the carbonyl group of Val 29 (0.25 Å closer to OH2') and Glu 30 (0.7 Å closer to OH2' and 0.9 Å closer to OH3'), and between the γ -phosphate group and the amide of Gly 60 and Thr 35 (both 0.25 Å closer). Tyr 32, which has been found to change its orientation drastically when going from the GDP-bound to the GTP-bound state^{29,34}, forms a strong hydrogen bond to one oxygen of the γ -phosphate. These tighter hydrogen bonds, together with the shielding of the phosphate part of the nucleotide-binding pocket from the solvent as a result of packing between the aromatic ring of Tyr 32 and the pyrrolidine ring of Pro 34 (Fig. 5c), account for the observation that RBD acts as a guanine-nucleotide dissociation inhibitor (GDI) on Ras and Rap¹³.

A water molecule, in a position equivalent to that of Wat 175 in the uncomplexed Ras-GppNHp²⁴, is found in the Rap-RBD complex. It is close (3.6 Å) to the γ -phosphorus and in a position suitable for nucleophilic attack. It is also able to make a hydrogen bond to one oxygen of the γ -phosphate, as found for Ras. This water molecule is also coordinated to the carbonyl of Thr 35 and to the hydroxyl group of Thr 61 (2.8 Å), the residue equivalent to Gln 61 in Ras and responsible for the reduced intrinsic GTPase activity of Rap¹⁶.

FIG. 2 Structure of the Rap-RBD complex. Ribbon representation: a, view perpendicular to the β -sheet shared by the two proteins; and b, 90° rotation around a horizontal axis in the plane of the paper. Rap, in yellow; RBD, red. The secondary structure elements of RBD as assigned by the program DSSP⁴⁸ are labelled. c, Stereo view of the $C\alpha$ trace of the complex oriented as in a; GppNHp is in red. Drawn with MOLSCRIPT⁴⁹.

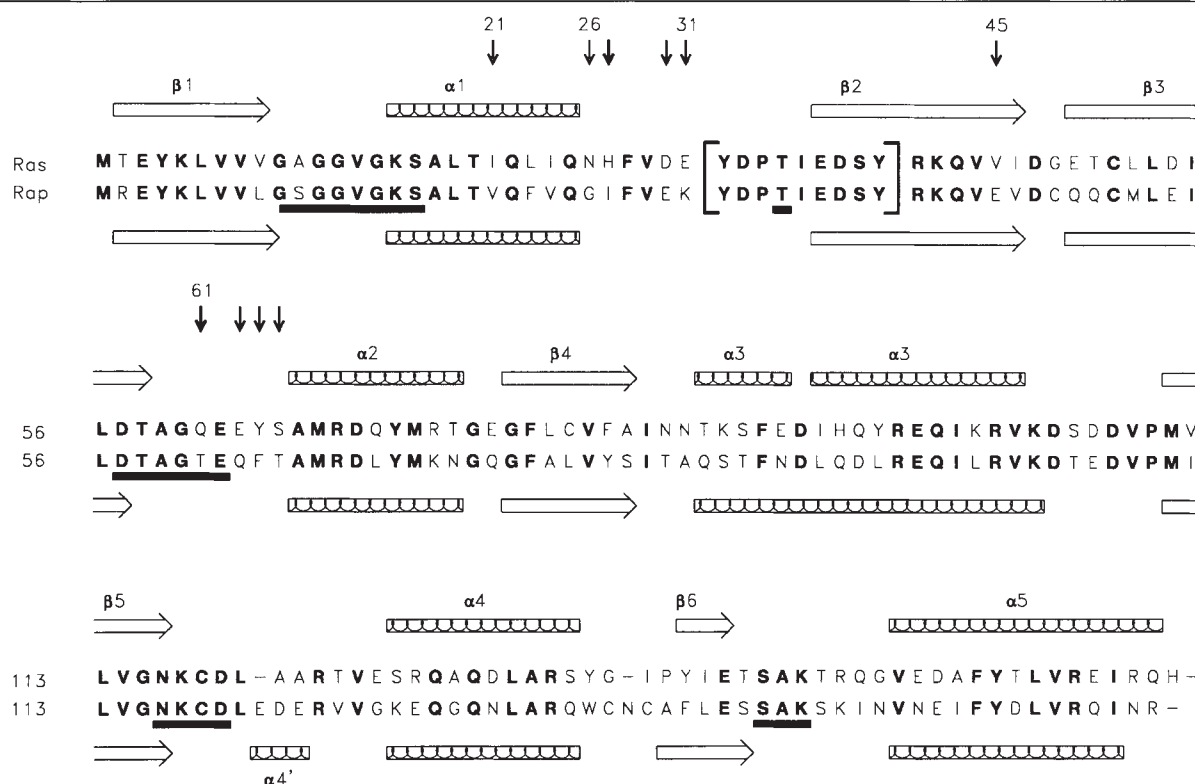
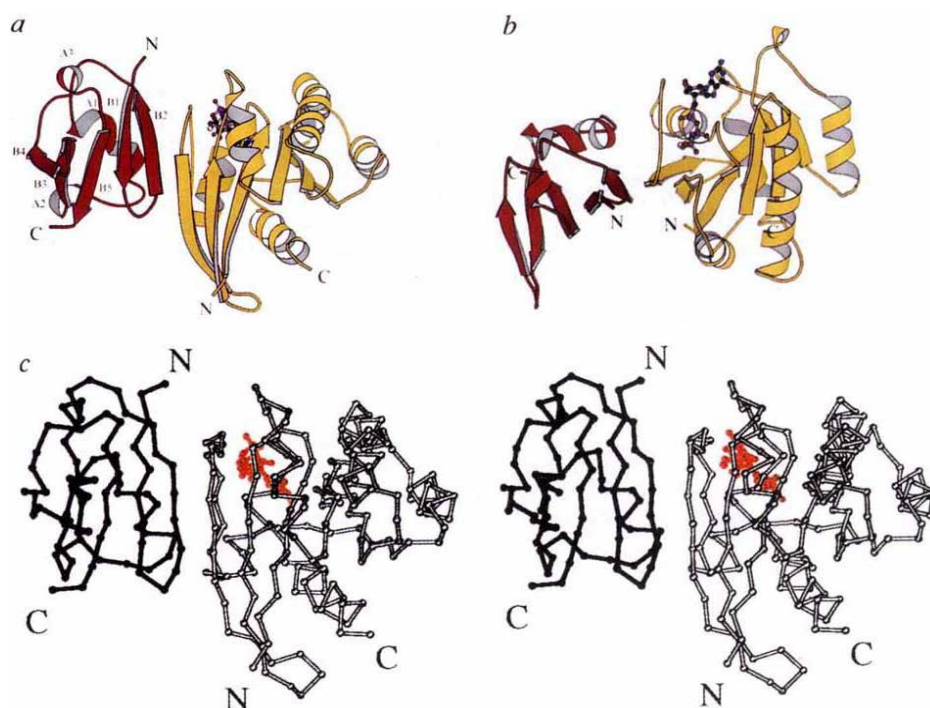


FIG. 3 Alignment and structure assignment for Ras and Rap. The alignment is based on the structure of human H-Ras(1–166) in the GppNHp form²⁴ and the structure of human truncated Rap1A (residues 1–167) in the complex with the RBD domain as shown here. Secondary structure elements were analysed using DSSP⁴⁸ and are shown above and below the sequences. The amino-acid residues

that differ between Ras and Rap, which may be responsible for their opposing biological functions^{15,32,33}, are indicated by arrows. Of these, only residues 21 and 27 are located in the protein–protein interface and are apparently involved in binding. Residues found in all GTP-binding proteins are underlined; identical effector residues 32–40 are boxed.

The Ras-binding domain

Figure 4a shows the primary structure alignment of several vertebrate and invertebrate RBD sequences, together with the secondary structure assignments as determined from the crystal structure. As suggested before, on the basis of heteronuclear

three-dimensional NMR spectroscopy of a similar fragment of Raf containing residues 55–132 (ref. 11), the topology of RBD is very similar to that of ubiquitin³⁵. The ubiquitin fold has been classified as a protein superfold, the ubiquitin α/β roll³⁶; the 2Fe–2S ferredoxin from green algae and the B1 domain of strep-

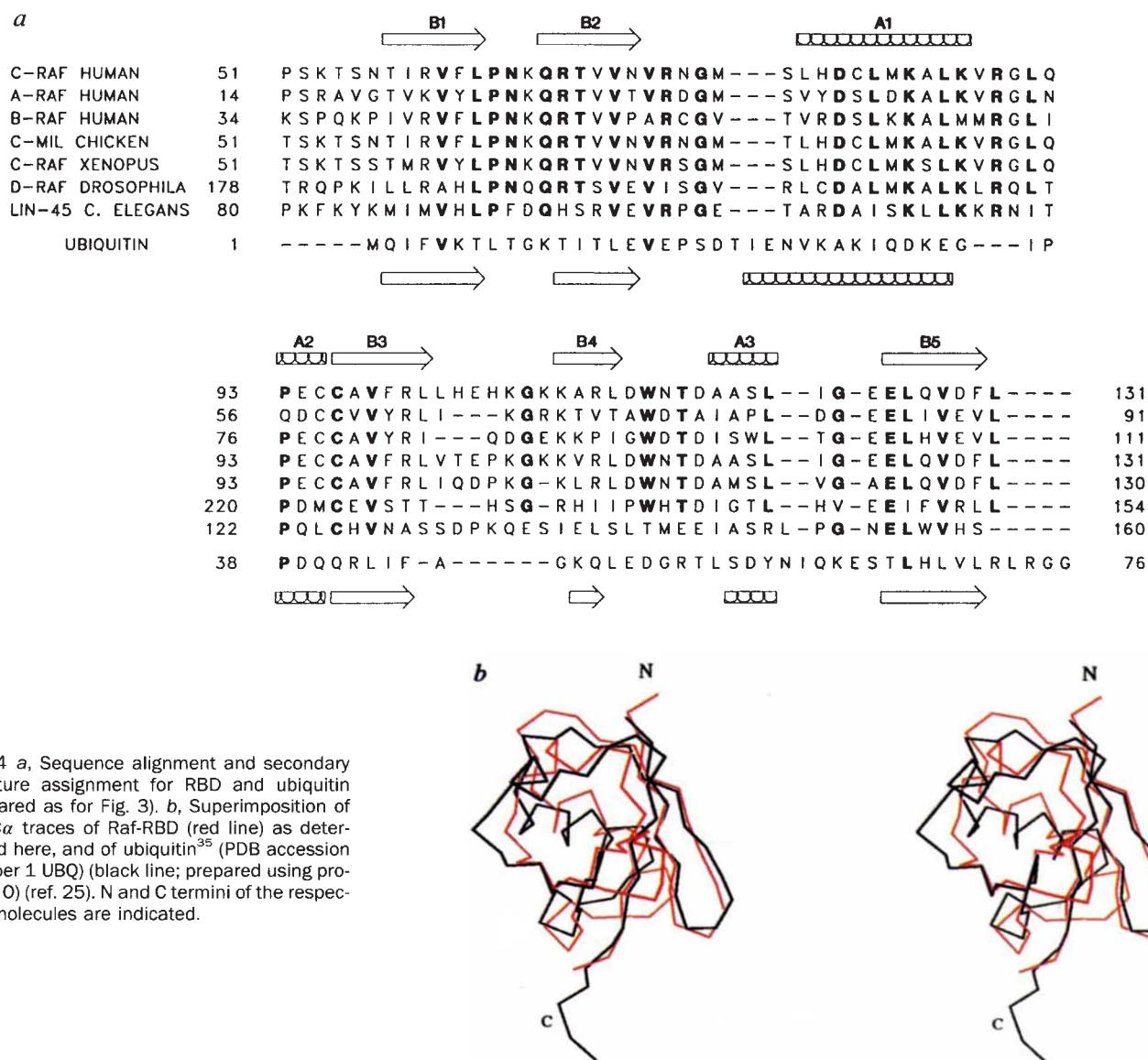


FIG. 4 *a*, Sequence alignment and secondary structure assignment for RBD and ubiquitin (prepared as for Fig. 3). *b*, Superimposition of the α traces of Raf-RBD (red line) as determined here, and of ubiquitin³⁵ (PDB accession number 1 UBQ) (black line; prepared using program O) (ref. 25). N and C termini of the respective molecules are indicated.

tococcal protein G, which is involved in immunoglobulin binding, have the same topology. These domains are unrelated by primary sequence (<25% sequence identity).

The RBD domain of Raf consists of a five-stranded mixed β -sheet (B1–B5) with an interrupted α -helix (A1) and two additional small 3_{10} -helices (A2 and A3). Superimposition of ubiquitin with the RBD domain (Fig. 4*b*) gives an average r.m.s. deviation of 1.7 Å over the superimposed α backbone. The largest deviations are for residues 73–76 in the loop connecting B2 with the α -helix (4–9 Å) and the loop connecting B4 and B5 (3–5 Å) and containing the 3_{10} helix A3. It is interesting to note that ubiquitin does not contain residues homologous to residues 88–90 of RBD, which include Arg 89 at the centre of the binding interface (see later). The disordered loop connecting B3 and B4, which is shorter in A- and B-Raf, also has no counterpart in the ubiquitin structure.

The complex

Figure 2 shows two different views of the Rap/RBD complex. It illustrates that the major interaction is between residues of the two antiparallel β -strands— β 2 from Rap and B2 from RBD—and that in addition only the C-terminal end of helix A1 is in contact with Rap (Fig. 5). The size of the buried interface, calculated by rolling a probe sphere of 1.4 Å to be 1,175 Å²

(NACCESS program written by S. Hubbard and J. Thornton), or 1,255 Å² (ref. 37) for the accessible surface, is in the range found for other protein–protein interfaces³⁸. A survey of the buried surface of several protein–protein complexes indicates that there is no correlation between the strength of the association of the complex and the size of the interface. The calculated shape complementarity (Sc) parameter for this complex is 0.76, a value close to that calculated for oligomeric proteins and protein/protein-inhibitor interfaces, which shows that there is good shape complementarity at the interface³⁹. Surprisingly, the binding interface is not close to the γ -phosphate or any other part of the nucleotide, as would be predicted from the GDI effect of RBD on Ras and Rap (ref. 13 and Fig. 5*c*).

GTP-binding proteins are generally involved in the transmission of signals or in triggering a transport process. This is accomplished by the interaction between the GTP-bound state of the protein and a target molecule, the effector. In the case of Ras, residues involved in effector interaction have originally been defined genetically and are localized in the so-called effector region^{32,33}, which comprises residues 32–40 and is identical in Ras and Rap (Fig. 3). Ras and Rap have opposing effects on cell transformation and both effects are disrupted by mutation of Asp 38 in the effector region; introduction of Ras residues 26, 27, 30, 31 and 45 into Rap is sufficient for transformation^{15,32,33}.

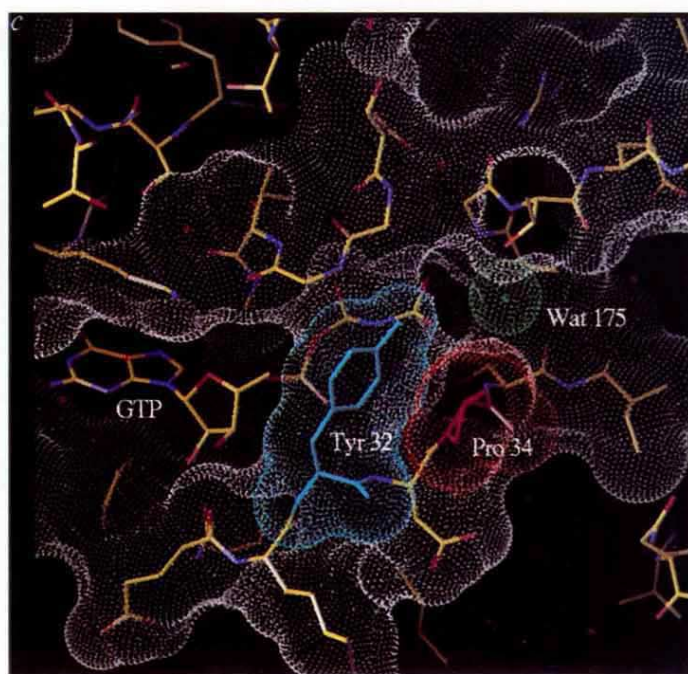
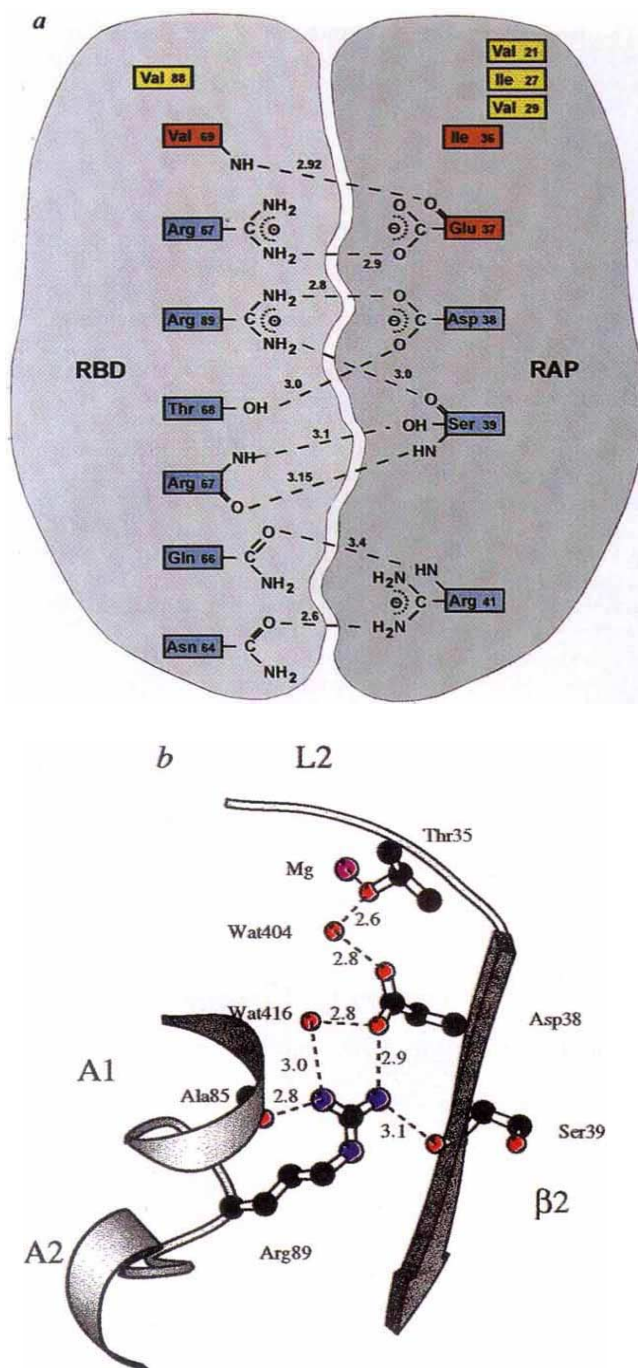


FIG. 5 Details of the interface between Rap and RBD and the GTP-binding site of Rap. *a*, Schematic view of the complex interface with hydrogen bonds (dashed lines) and their distances in Å. Residues involved in polar interactions are coloured in blue, those in hydrophobic contacts are in orange (contact area 1) and in yellow (contact area 2). In addition to the direct contacts shown here there are water-mediated contacts between Ile 36 and Val 69, Asp 38 and Arg 89 (*b*), Ser 39 and Lys 65, Glu 59 and Arg 67 and Arg 41 and Gln 66 of Rap and RBD respectively. *b*, View of the network of hydrogen bonds between residues Asp 38 from Rap and Arg 89 from RBD. *c*, Accessible surface model of the GTP-binding site of Rap in the Rap-RBD structure, calculated by the Connolly algorithm³⁷. Blue and orange are the surfaces of Tyr 32 and Pro 34, respectively, which are packed against each other and cover the phosphate-binding area. A water molecule (green); marked as Wat175, which is equivalent to Wat175 in the Ras-GppNHp structure²⁴ can access the γ-phosphate through a small open channel just large enough for water to enter and P_i to leave.

31, which may mark the difference between Ras and Rap^{32,33}, are close to the protein-protein interface but make no direct contact, although Lys 48 is close to these residues.

Implications for the signalling complex

As most of the residues involved in complex formation, apart from Val 21 and Ile 27 in one of the hydrophobic pockets of the interface, are conserved between Ras and Rap, and the structures of Ras and Rap are nearly identical, we can assume that the structure of the complex between RBD and Ras-GppNHp is identical or nearly identical (preliminary crystallographic analysis of that complex does in fact support this). The three-dimensional structure of the complex can now be related to the numerous mutational studies on Ras^{32,33}, one being the Ile-36 mutant, which is weakly transforming in viral Ras when substituted by valine or leucine but not when substituted by alanine. Although there are no biochemical data available on the Rap-Raf interaction for such mutants, the structure indicates that the hydrophobic interaction of Ile 36 is interrupted by alanine, but that mutation to Val/Leu can be accommodated by the protein-protein interface. The most studied mutation to affect Ras-effector interaction involves substitution of alanine for aspartate at residue 38 (D38A), which reduces the affinity for RBD 72-fold¹³, and even the conservative D38E mutation abolishes the biologi-

The structural details of the interaction between Rap and RBD are shown schematically in Fig. 5*a*. The interaction is mediated mostly by hydrogen bonds and polar interactions from charged residues, from both main-chain and side-chain polar groups, and only a few hydrophobic contacts. As expected, residues from the more closely defined effector region such as Asp 33, Ile 36, Glu 37, Asp 38, Ser 39 are in close contact with RBD. From the ribbon representation in Fig. 2 and from the similar distances between consecutive Cα positions of four antiparallel strands, it is likely that the two proteins form a continuous antiparallel β-sheet between B1-B2 from RBD and β2-β3 from Rap (Fig. 1*b*). The contacts between the strands are made by two main-chain and many side-chain interactions; there are also contacts between the two proteins mediated by four water molecules and residues Ile 36, Glu 37, Asp 38, Ser 39, Glu 54 and Arg 41 from Rap and residues Val 69, Lys 65, Arg 67, Asn 64 and Arg 89 from RBD (not shown). Residues 26, 30 and

cal activity of Ras⁴⁰. The structure in Fig. 5a,b shows that Asp 38 makes two direct hydrogen bonds to the hydroxyl of Thr 68 and the side chain of Arg 89 from RBD, and water-mediated hydrogen bonds to residues Val 69 (not shown) and Arg 89 from RBD, and to the hydroxyl of Thr 35 of Rap, which is a ligand for Mg²⁺.

Much less is known about residues on Raf that are important for interaction with Ras. The mutation R89L, originally identified as a mutation that interrupts signalling by the *Drosophila* homologue of Raf kinase, is unable to bind to Ras-GTP *in vitro*⁴¹. As already mentioned, this residue forms a direct and a water-mediated interaction with Asp 38 and a direct hydrogen bond to the carbonyl of Ser 39 (Fig. 5b). Contact epitope scanning and alanine scanning mutagenesis have been used to identify the regions 88–91 and 101–103 as important for Ras–Raf interaction⁴². For a detailed analysis of the binding between Ras/Rap and Raf it will be necessary to mutate the residues involved and measure the effects on the properties of the complex.

The structure of the Rap–RBD complex accounts for many, but not all, of the mutational effects of Ras. For example, the V45E mutation in viral Ras is not transforming, even though the structure of the complex shows that Glu 45 in Ras is far removed from the interaction surface of RBD. Also, the P34R mutation is potentially oncogenic in wild-type Ras, which the structure of the complex cannot readily explain. It is possible that mutations such as V45E, P34G and T35A destabilize the structure and that other elements of the regulatory region of Raf are contributing to the Ras–Raf or Rap–Raf interaction. It is known that intragenic mutations in the zinc-binding region of D-Raf that are outside the RBD domain suppress a signalling deficiency in the R89L mutation of D-Raf in the Sevenless pathway⁴³ and that the E37G mutation in Ras interrupts the interaction with Raf1, which can be predicted from the structure, but this defect can be restored by mutations in the Ser/Thr-rich CR2 region of Raf (the second of the highly conserved regions of Raf-homologous proteins)⁴⁴. Other effectors may interact with different surface residues on Ras: for example, the E37G and T35S Ras mutations influence the interactions with different effectors⁴⁴.

GTPase cycle and conformational change

Wild-type Ras cycles between the GDP-bound and GTP-bound state, and the GTPase reaction is the timer for the interaction between Ras-GTP and the effector such as the Raf protein kin-

ase. The structure in Fig. 5c shows that a water molecule is situated close to the γ -phosphorus of GppNHP in the Rap–RBD complex in a similar position to Wat 175 in Ras–GppNHP alone²⁴ and, just as in Ras, it can donate its hydrogen to one of the oxygens of the γ -phosphate as a trigger for the GTP hydrolysis reaction. After hydrolysis of the Ras·GTP–RBD complex, inorganic phosphate is released and Ras·GDP–RBD is formed¹³. The structure of the complex shows (Fig. 5c) that although the base and the phosphates, but not the ribose of GTP in the Rap–Raf complex, are shielded from solvent, there is a small channel between the γ -phosphorus and the solvent that is just large enough to allow a water molecule to enter and inorganic phosphate to leave.

As the GDP-bound form of Ras has a 1,300-fold lower affinity for RBD it should dissociate immediately¹³. We cannot explain how the conformation changes from the high-affinity to the low-affinity complex from the structure of the Rap–RBD complex, but it must be triggered by releasing Thr 35 and Tyr 32 from interaction with the γ -phosphate and be transmitted to the switch-I region of Ras/Rap, leading to changes in the interface of the Ras/Rap–Raf complex. One such residue is Asp 38, which is connected to Thr 35 through a water molecule; disrupting this connection may loosen the Asp 38–Arg 89 interaction. The switch-II region comprising residues in loop L4/ α 2 of Ras/Rap is not directly involved in the interaction with RBD, even though its structure is altered in the conformational change²⁹ and it has been implicated in the interaction with effectors.

Outlook

We have identified for the first time the features of the interaction between a GTP-binding protein and the binding domain of its possible effector. The protein uses its single antiparallel β -strand at the edge of the molecule to accomplish binding, which may be an interaction that can be used and modified by many other GTP-binding proteins. It will be interesting to see whether Ras interacts in the same way with other regulators and effectors such as GAP, Raf-GEF or PI(3)K and whether other GTP-binding proteins like Rho, Rab and Ran make use of similar interactions.

The interface of the Rap–RBD complex, and by inference, of the Ras–RBD, appears to be small enough to allow inhibitors of this interaction to be designed as anti-Ras drugs, providing the Ras–Raf interaction is confirmed as the Achilles heel⁴⁵ of Ras-dependent transformation. □

Received 13 March; accepted 5 May 1995.

1. Daum, G., Eisenmann-Tappe, I., Fries, H.-W., Troppmair, J. & Rapp, U. R. *Trends biochem. Sci.* **19**, 474–480 (1994).
2. Burgering, B. M. T. & Bos, J. L. *Trends biochem. Sci.* **20**, 18–22 (1995).
3. Koide, H., Satoh, T., Nakafuku, M. & Kaziro, Y. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8683–8686 (1993).
4. Moodie, S. A., Willumsen, B. M., Weber, M. J. & Wolfman, A. *Science* **260**, 1658–1661 (1993).
5. van Aelst, L., Barr, M., Marcus, S., Polverino, A. & Wigler, M. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6213–6217 (1993).
6. Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. *Cell* **74**, 205–214 (1993).
7. Zhang, X. F. et al. *Nature* **364**, 308–313 (1993).
8. Warne, P. H., Rodriguez Viciano, P. & Downward, J. *Nature* **364**, 352–355 (1993).
9. Leever, S. J., Paterson, H. F. & Marshall, C. J. *Nature* **369**, 411–414 (1994).
10. Stokoe, D., Macdonald, S. G., Cadwallader, K., Symons, M. & Hancock, J. F. *Science* **264**, 1463–1467 (1994).
11. Emerson, S. D. et al. *Biochemistry* **33**, 7745–7752 (1994).
12. Chuang, E. et al. *Molec. cell. biol.* **14**, 5318–5325 (1994).
13. Herrmann, C., Martin, G. A. & Wittinghofer, A. *J. biol. Chem.* **270**, 2901–2905 (1995).
14. Pizon, V., Chardin, P., Leresoy, I., Olofsson, B. & Tavittian, A. *Oncogene* **3**, 201–204 (1988).
15. Bokoch, G. in *GTPases in G-protein Signaling*, Vol. 108/1 (eds Dickey, B. F. & Birnbaumer, L.) 377–393 (Springer, Berlin and Heidelberg, 1993).
16. Frech, M. et al. *Science* **249**, 169–171 (1990).
17. Hata, Y. et al. *J. biol. Chem.* **265**, 7104–7107 (1990).
18. Hofer, F., Fields, S., Schneider, C. & Martin, G. S. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11089–11093 (1994).
19. Kikuchi, A., Demo, S. D., Ye, Z.-H., Chen, Y.-W. & Williams, L. T. *Molec. cell. Biol.* **14**, 7483–7491 (1994).
20. Spaargaren, M. & Bishoff, J. R. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12609–12613 (1994).
21. Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. *Cell* **58**, 77–84 (1989).
22. Hariharan, I. K., Carthew, R. W. & Rubin, G. M. *Cell* **67**, 717–722 (1991).
23. Cook, S., Rubinfeld, B., Albert, I. & McCormick, F. *EMBO J.* **12**, 3475–3485 (1993).
24. Pai, E. F. et al. *EMBO J.* **9**, 2351–2359 (1990).
25. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
26. Brunger, T. A. *XPLOR* version 3.1, Yale Univ. (1993).

27. Read, R. J. *Acta crystallogr.* **A42**, 140–149 (1986).
28. Brunger, T. A. *Nature* **355**, 472–475 (1992).
29. Milburn, M. V. et al. *Science* **247**, 939–945 (1990).
30. Amor, J. C., Harrison, D. H., Kahn, R. A. & Ringe, D. *Nature* **372**, 704–708 (1994).
31. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. *Nature* **374**, 378–381 (1995).
32. Marshall, M. S. *Trends biochem. Sci.* **18**, 250–254 (1993).
33. Polakis, P. & McCormick, F. *J. biol. Chem.* **268**, 9157–9160 (1993).
34. Schlichting, I. et al. *Nature* **345**, 309–315 (1990).
35. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. *J. molec. Biol.* **194**, 531–544 (1987).
36. Orengo, C. A., Jones, D. T. & Thornton, J. M. *Nature* **372**, 631–634 (1994).
37. Connolly, M. L. *J. appl. Crystallogr.* **16**, 548–558 (1983).
38. Janin, J. & Chothia, C. *J. biol. Chem.* **265**, 16027–16030 (1990).
39. Lawrence, M. C. & Colman, P. M. *J. molec. Biol.* **234**, 946–950 (1993).
40. Calés, C., Hancock, J. F., Marshall, C. J. & Hall, A. *Nature* **332**, 548–551 (1988).
41. Fabian, J. R., Vojtek, A. B., Cooper, A. & Morrison, D. K. *Proc. natn. Acad. Sci. U.S.A.* **91**, 5982–5986 (1994).
42. Barnard, D. et al. *Oncogene* **10**, 1283–1290 (1995).
43. Lu, X., Melnick, M. B., Hsu, J.-C. & Perrimon, N. *EMBO J.* **13**, 2592–2599 (1994).
44. White, M. et al. *Cell* **80**, 1–20 (1995).
45. Gibbs, J. B. & Oliff, A. *Cell* **79**, 193–198 (1994).
46. Kabsch, W. *J. appl. Crystallogr.* **26**, 795–800 (1993).
47. Navaza, J. *Acta crystallogr.* **A50**, 157–163 (1994).
48. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
49. Kraulis, P. J. *J. appl. Crystallogr.* **24**, 946–950 (1991).

ACKNOWLEDGEMENTS. We thank V. Pizon for the clone of truncated Rap1A; K. Scheffzek, D. Madden, B. Tilton (Bayer) and W. Kabsch for discussion and for technical advice; F. Schmitz and I. Vetter for help with graphical presentation; R. Schebaum for secretarial assistance; W. K. and K. C. Holmes for the use of facilities at the MPI für medizinische Forschung (Heidelberg); and K. Wilson and C. Betzel for help with synchrotron data collection. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) and the European Community. Coordinates will be deposited in the Brookhaven Protein Databank, where they will be held for one year.

Evidence for stripe correlations of spins and holes in copper oxide superconductors

J. M. Tranquada*, B. J. Sternlieb†, J. D. Axe*, Y. Nakamura† & S. Uchida†

* Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA

† Superconductivity Research Course, The University of Tokyo, Yayoi 2-11-16, Bunkyo-ku, Tokyo 113, Japan

ONE of the long-standing mysteries associated with the high-temperature copper oxide superconductors concerns the anomalous suppression¹ of superconductivity in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$ (and certain related compounds) when the hole concentration x is near $\frac{1}{8}$. Here we examine the possibility that this effect is related to dynamical two-dimensional spin correlations, incommensurate with the crystal lattice, that have been observed in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ by neutron scattering²⁻⁴. A possible explanation for the incommensurability involves a coupled, dynamical modulation of spin and charge in which antiferromagnetic 'stripes' of copper spins are separated by periodically spaced domain walls to which the holes segregate⁵⁻⁹. An ordered stripe phase of this type has recently been observed in hole-doped La_2NiO_4 (refs 10-12). We present evidence from neutron diffraction that in the copper oxide material $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$, with $x=0.12$, a static analogue of the dynamical stripe phase is present, and is associated with an anomalous suppression of superconductivity^{13,14}. Our results thus provide an explanation of the '1/8' conundrum, and also support the suggestion¹⁵ that spatial modulations of spin and charge density are related to superconductivity in the copper oxides.

To clarify the nature of the correlations with which we are concerned, we begin by describing the stripe modulation that has been observed¹⁰⁻¹² in nickel oxide analogues of the copper oxides, La_2NiO_4 ,^{12,5} and $\text{La}_{1.8}\text{Sr}_{0.2}\text{NiO}_4$. The spatial correlations of spins and holes in an NiO_2 plane are illustrated schematically in Fig. 1a. The magnetic moments on Ni atoms are locally antiferromagnetic. For a two-dimensional square lattice, antiferromagnetic order is characterized by the wave vector $(\frac{1}{2}, \frac{1}{2})$ (where we specify coordinates in terms of reciprocal lattice units, $2\pi/a$). In the case of the modulated spin structure, the characteristic wave vectors are $(\frac{1}{2} \pm \epsilon, \frac{1}{2} \pm \epsilon)$ and $(-\frac{1}{2} \pm \epsilon, \frac{1}{2} \pm \epsilon)$, where, for the case shown, $\epsilon = \frac{1}{8}$. One might imagine that higher harmonics should be important for describing the spin structure; however, even for the extreme case of sharp domain walls shown in the figure, Fourier analysis shows that the only other distinct harmonic is the third [wave vectors $(\frac{1}{2} \pm 3\epsilon, \frac{1}{2} \pm 3\epsilon)$], and its intensity is only 3% of the first harmonic. The experimentally observed third-harmonic intensities are somewhat smaller than this, indicating a more sinusoidal modulation of the domains.

The period of the charge modulation is only half that of the spins, and as a result corresponding diffraction peaks appear at second-harmonic positions. Of course, the charge distribution cannot be detected directly by neutrons; instead, we measure the modulation of atomic positions associated with the charge modulation (a charge-density wave). The characteristic three-dimensional wave vectors for the atomic displacements are $(h, k, l) = (\pm 2\epsilon, 0, 1)$. The requirement that $l=1$ is related to the presence of two NiO_2 layers per unit cell, and indicates that the displacement pattern in one layer is exactly opposite to that in the layer just above (and below) it. This result is consistent with a staggering of the hole stripes from one layer to the next, as one would expect because of Coulomb repulsion. A quantitative analysis¹² of the observed superlattice intensities measured in the

oxygen-doped crystal has verified that the positions of the Ni and O atoms in the planes are sinusoidally modulated along the direction perpendicular to the stripes, as expected for a charge-density wave. Although it is not possible to infer the charge distribution directly from the atomic displacements, the observations are generally consistent with recent multiband Hubbard model calculations by Zaanen and Littlewood¹⁶.

The spin and hole ordering in the nickel oxides described above occurs for temperatures near 100 K. A striking feature of the ordering is its cooperative nature: the charge and spin order parameters show similar temperature dependences^{10,11}. The ordered magnetic moments are found to be quite large, with a maximum moment (in the oxygen-doped crystal¹²) of approximately 85% of that found in the undoped antiferromagnetic-insulator phase. Because a hole has its own magnetic moment which will tend to couple to a Ni moment in a way that reduces the ordered value, the large observed moments indicate that the hole density is strongly modulated, as assumed in the domain-wall/stripe picture. The localization of the holes in domain walls increases the kinetic energy of the holes, while decreasing the potential energy of the Ni spins. The competition between these energies drives the density modulation.

What has the above to do with the copper oxides? After all, the nickel oxides remain insulating up to rather high dopant concentrations. One reason for considering a connection comes from mean-field analyses of the single-band Hubbard model⁵⁻⁷, a simple model believed to contain the basic physics of a CuO_2 plane. If electron-electron interactions are assumed strong and spatial modulation of the spin and charge densities is allowed, stripe phases involving charged domain walls are commonly found. Another reason is the observation of incommensurate magnetic peaks in inelastic neutron scattering studies²⁻⁴ of super-

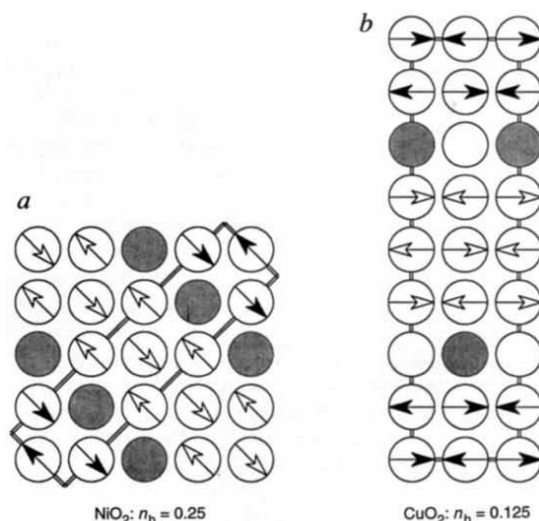


FIG. 1 a, Idealized diagram of the spin and charge stripe pattern within a NiO_2 plane observed in hole-doped La_2NiO_4 with a hole density of $n_h = \frac{1}{4}$. b, Hypothesized stripe pattern in a CuO_2 plane of hole-doped La_2CuO_4 with $n_h = \frac{1}{8}$. In both, only the metal atoms are represented; the oxygen atoms, which surround the metal sites in a square planar array (as shown in Fig. 2), have been omitted. Arrows indicate the orientation of magnetic moments on metal atoms, which are locally antiparallel; the spin direction rotates by 180° (relative to a simple antiferromagnetic structure) on crossing a domain wall, as emphasized by the change in filling of the arrow heads. The doubled lines outline the magnetic unit cell in each case. Holes are located at the anti-phase domain boundaries, which are indicated by the rows of circles without arrows. A filled circle denotes the presence of one hole, centred on a metal site. For the present analysis, the hole density is assumed to be uniform along a domain wall; the ordering of holes along stripes suggested in b has not been tested experimentally, but serves as a reminder that the hole per Cu ratio is $\frac{1}{2}$.

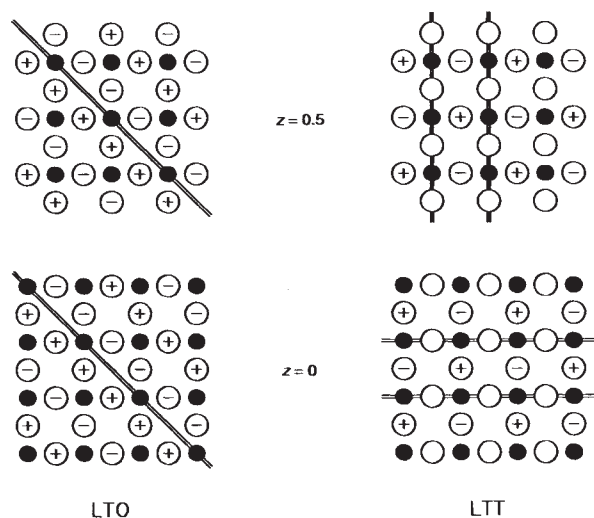


FIG. 2 Displacement patterns within the two CuO_2 layers of a unit cell for the LTO and LTT structures. Open (solid) circles represent oxygen (copper) atoms. The oxygen atoms are displaced out of the plane (+ or -) by local rotations of square planar CuO_4 units about tilt axes shown by the doubled lines. In the LTT structure the tilt axes rotate by 90° from $z=0$ to $z=0.5$, where z is the height along the c axis in lattice units.

conducting $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. Those peaks are characterized by the two-dimensional wave vectors $(\frac{1}{2} \pm \varepsilon, \frac{1}{2})$ and $(\frac{1}{2}, \frac{1}{2} \pm \varepsilon)$ with $\varepsilon \approx x$ (previous authors²⁻⁴ have used the parameter $\delta = 2\varepsilon$ to describe the splitting). Although no static component is observed, those peaks are consistent with instantaneous correlations of the type shown in Fig. 1b, with horizontal, rather than diagonal, domain walls. Besides the orientation of the stripes, another difference from the nickel oxides is the hole concentration in the domain walls, corresponding to only one hole for every two Cu sites. This constraint conflicts with the strong-interaction mean-field analyses, which predict⁷ one hole per site ($\varepsilon = \frac{1}{2}x$), and led initially to rejection² of the stripe interpretation. (A related problem is that the domain wall solutions are insulating⁷.) An alternative approach¹⁷⁻¹⁹ that assumes weak correlations leads to incommensurate spin modulations with a period determined by the shape and size of the Fermi surface. Suitable tuning of the band structure can yield agreement with the neutron observations¹⁸⁻¹⁹; however, the weak-correlation picture would not be consistent with a strong modulation of the charge and spin densities, such as the existence of a static segregation of holes into domain walls in lightly-doped La_2CuO_4 , as inferred from a study of bulk susceptibility²⁰.

If one believes that charge and spin stripe correlations exist in the copper oxides, where might one look for corroborating evidence? It has been suggested (ref. 15 and J. Zaanen, personal communication) that partial substitution of Zn for Cu, which is known to suppress superconductivity, might also pin domain walls, leading to static correlations; however, disorder would tend to make them difficult to detect. Instead, we have chosen to investigate whether stripe order might be associated with the anomalous suppression of superconductivity at a dopant concentration near $x = \frac{1}{8}$ first observed¹ in $\text{La}_{2-x}\text{Ba}_x\text{CuO}_4$. Studies of this and related copper-oxide systems have established that the suppression phenomenon requires, in addition to the unique hole concentration^{13,21,22}, a distortion of the lattice from the usual low-temperature orthorhombic (LTO) to the low-temperature tetragonal (LTT) structure^{13,23,24}. The LTT phase can be stabilized in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ by partial substitution of Nd for La ¹³, with the width (in x) of the suppression anomaly increasing at

higher Nd concentrations²⁴. The fact that the suppression occurs when the hole concentration is close to a simple fraction suggests that a commensurability effect is involved, and commensurate pinning of a striped phase could yield long-range order. Also suggestive is the observation of static magnetic correlations in $\text{La}_{1.875}\text{Ba}_{0.125}\text{CuO}_4$ by muon spin-rotation measurements^{25,26}.

A comparison of the LTO and LTT structures (Fig. 2) makes it clear why the latter phase might be especially favourable for pinning a horizontal stripe phase. The atomic displacements in the LTO structure form a diagonal pattern, whereas the pattern of displacements is horizontal (or vertical) in the LTT case. Thus horizontal charge stripes would tend to be pinned by the lattice potential in the LTT, but not in the LTO, phase.

Our recent neutron scattering study confirms this conjecture. A piece ($\sim 0.1 \text{ cm}^3$) of a $\text{La}_{1.48}\text{Nd}_{0.4}\text{Sr}_{0.12}\text{CuO}_4$ single crystal (the transport properties of which were reported previously¹⁴) was studied by neutron diffraction using triple-axis spectrometers at the High Flux Beam Reactor located at Brookhaven National Laboratory. The crystal was initially mounted in a cryostat with the [001] axis vertical, so that the $(hk0)$ zone of reciprocal space could be probed (Fig. 3a). In a scan along $(\frac{1}{2}, \frac{1}{2} + q, 0)$ at a temperature of 11 K (Fig. 3b), peaks are clearly observed at $|q| = \varepsilon = 0.12$. (The neutron wavelength had to be tuned to minimize multiple scattering at $q=0$, a small remnant of which is visible in the figure.) These peaks are very close to the commensurate positions ($\varepsilon = \frac{1}{8}$) that we expect for magnetic reflections if the Cu spins order in the horizontal stripe pattern shown in Fig. 1b. Also, ε is consistent with the values observed in the inelastic magnetic scattering studies²⁻⁴ of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. Third-harmonic peaks are not and should not be detectable in the present experiment (because of the signal-to-noise level), so we cannot distinguish between relatively sharp domain walls and broader density-wave modulations.

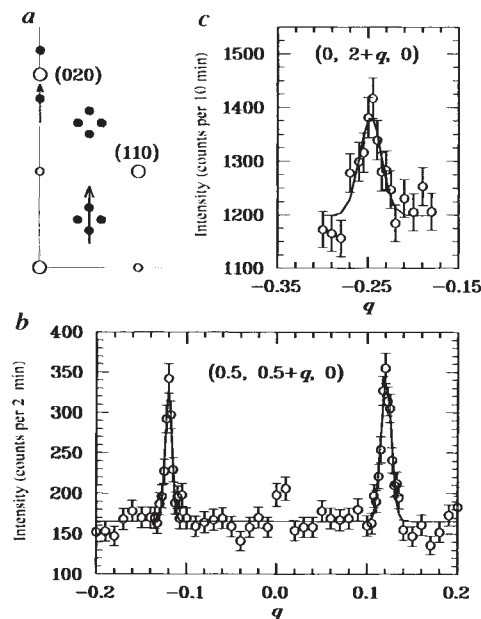


FIG. 3 Scans of superlattice peaks consistent with the proposed spin and charge stripes, in $\text{La}_{1.48}\text{Nd}_{0.4}\text{Sr}_{0.12}\text{CuO}_4$ at 11 K. a, Diagram of the $(hk0)$ zone of reciprocal space. Open circles indicate locations of Bragg peaks for the LTT structure; solid circles denote spin- and charge-ordering superlattice peaks. Arrows indicate the regions scanned. b, Scan along $(\frac{1}{2}, \frac{1}{2} + q, 0)$ through the $(\frac{1}{2}, \frac{1}{2} + \varepsilon, 0)$ peaks measured with a neutron energy of 13.9 meV. The small peak width indicates that the in-plane correlation length is greater than 150 Å. c, Scan along $(0, 2 + q, 0)$ through the $(0, 2 - 2\varepsilon, 0)$ peak using 14.7-meV neutrons. The lines in b and c are the result of least-squares fits to gaussian peak shapes plus a flat background.

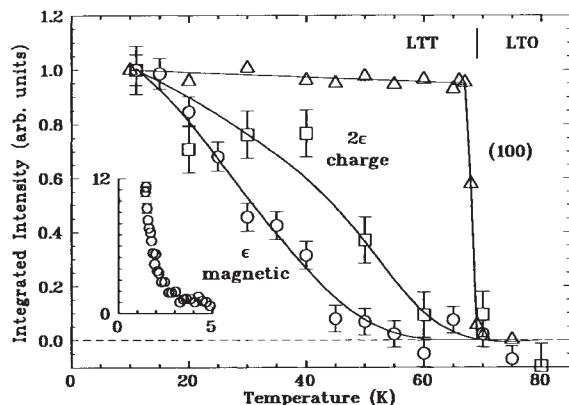


FIG. 4 Temperature dependence of superlattice peak intensities, normalized at 11 K, measured in $\text{La}_{1.48}\text{Nd}_{0.4}\text{Sr}_{0.12}\text{CuO}_4$. Results shown for the magnetic $(\frac{1}{2}, \frac{1}{2} - \epsilon, 0)$ (circles), charge-related $(0, 2 - 2\epsilon, 0)$ (squares) and $(1, 0, 0)$ (triangles). The $(1, 0, 0)$ peak is allowed in LTT but not LTO. The lines are guides to the eye. Inset, temperature dependence of the magnetic peak intensity below 5 K.

The search for nuclear superlattice peaks associated with the hole stripes required further thought. We had initially expected, by analogy with the nickel oxides, that the characteristic wave vectors would be $(\pm 2\epsilon, 0, 1)$, with $\epsilon = \frac{1}{8}$; however, a search of such positions gave a negative result. Examination of Fig. 2 reveals why. If the stripes are truly pinned by the atomic displacement pattern within a layer, then they must rotate by 90° from one layer to the next to follow the LTT distortion pattern. Because of the degeneracy in positioning the spin and charge stripes within a plane, and a weak interaction between neighbouring orthogonal stripe layers, we should not expect the stripe pattern to exhibit strong correlations along the c axis. As a result, the scattering from the stripes should be essentially two-dimensional in character, with the modulation wave vectors $(\pm 2\epsilon, 0, l)$ for the vertical stripe layers and $(0, \pm 2\epsilon, l)$ for the horizontal stripe layers. After working through these arguments, we were able to locate the $(0, 2 - 2\epsilon, 0)$ peak shown in Fig. 3c. The intensity of the observed peak is $\sim 10^{-6}$ times that of the strong (020) nuclear reflection.

These results led us to reinvestigate the magnetic scattering. In the process, we discovered a strong increase in the magnetic peak intensity below 3 K (Fig. 4 inset) due to ordering of the Nd moments and the development of very weak correlations between next-nearest-neighbour layers (nearest-neighbour layers remain uncorrelated), as determined from an analysis of the intensity variation of the magnetic scattering as a function of the momentum transfer perpendicular to the planes. The data and analysis will be presented elsewhere.

The temperature dependences of the magnetic (ϵ) and charge (2ϵ) peaks are plotted in Fig. 4. Although the magnetic peaks decrease faster than the charge peaks, both types of order disappear at or before the LTT–LTO structural transition at 70 K is reached. The disappearance of the static order either at or before reaching the LTT–LTO transition is consistent with our picture,

as the LTT structure favours, but is not induced by, the stripe order. The difference in the temperature dependences of the static spin and charge order parameters is similar to that observed¹¹ in $\text{La}_{1.8}\text{Sr}_{0.2}\text{NiO}_4$.

Although further studies are needed to verify our interpretation of the second-harmonic peaks, both the similarities to and differences from the nickel oxides make a strong case for the existence of static spin and charge stripe order in our copper-oxide crystal. Accepting this point, it is then reasonable to assume that dynamical stripe correlations exist in the LTO phase. Pinning of the stripes in the LTT phase is associated with an increase in resistivity¹⁴ and the suppression of superconductivity.

The implications of our results go well beyond resolution of the long-standing ‘ $\frac{1}{8}$ problem’^{1,13,21–24}. The results appear to be consistent with the ideas of frustrated phase separation discussed by Emery and Kivelson^{8,9,15} in which the long-range part of the Coulomb force (which is assumed to be screened in the Hubbard model) plays an important role. They have suggested that correlated fluctuations of hole-rich and hole-poor regions within CuO_2 planes might explain many of the anomalous normal-state properties of the layered copper oxides. Further theoretical work would be required to give a quantitative explanation for the observed stripe correlations, and to understand whether and how they may be related to superconductivity. Experimentally, it is of interest to study how such charge and spin modulations might be realized in other copper-oxide families. For example, the inelastic magnetic scattering in metallic $\text{YBa}_2\text{Cu}_3\text{O}_{6+x}$ is found to be commensurate rather than incommensurate; nevertheless, evidence for dynamically correlated antiferromagnetic clusters with a characteristic size has been reported in a recent neutron scattering study²⁷ of $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. In any case, our evidence for spin and charge stripe correlations suggests their importance in the copper oxides and a connection with superconductivity. □

Received 1 March; accepted 19 May 1995.

1. Moodenbaugh, A. R., Xu, Y., Suenaga, M., Folkerts, T. J. & Shelton, R. N. *Phys. Rev. B* **38**, 4596–4600 (1988).
2. Cheong, S. W. et al. *Phys. Rev. Lett.* **67**, 1791–1794 (1991).
3. Mason, T. E., Aeppli, G. & Mook, H. A. *Phys. Rev. Lett.* **68**, 1414–1417 (1992).
4. Thurston, T. R. et al. *Phys. Rev. B* **46**, 9128–9131 (1992).
5. Schulz, H. J. *J. Phys. (Paris)* **50**, 2833–2849 (1989).
6. Zaanen, J. & Gunnarsson, O. *Phys. Rev. B* **40**, 7391–7394 (1989).
7. Poilblanc, D. & Rice, T. M. *Phys. Rev. B* **39**, 9749–9752 (1989).
8. Emery, V. J. & Kivelson, S. A. *Physica C* **235–240**, 189–192 (1994).
9. Kivelson, S. A. & Emery, V. J. in *Strongly Correlated Electronic Materials: The Los Alamos Symposium 1993* (eds. Bedell, K. S. et al.) 619–656 (Addison-Wesley, Reading, MA, 1994).
10. Tranquada, J. M., Buttrey, D. J., Sachan, V. & Lorenzo, J. E. *Phys. Rev. Lett.* **73**, 1003–1006 (1994).
11. Sachan, V., Buttrey, D. J., Tranquada, J. M., Lorenzo, J. E. & Shirane, G. *Phys. Rev. B* **51**, 12742–12746 (1995).
12. Tranquada, J. M., Lorenzo, J. E., Buttrey, D. J. & Sachan, V. *Phys. Rev. B* (in the press).
13. Crawford, M. K. et al. *Phys. Rev. B* **44**, 7749–7752 (1991).
14. Nakamura, Y. & Uchida, S. *Phys. Rev. B* **46**, 5841–5844 (1992).
15. Emery, V. J. & Kivelson, S. A. *Physica C* **209**, 597–621 (1993).

16. Zaanen, J. & Littlewood, P. B. *Phys. Rev. B* **50**, 7222–7225 (1994).
17. Bulut, N., Hone, D., Scalapino, D. J. & Bickers, N. E. *Phys. Rev. Lett.* **64**, 2723–2726 (1990).
18. Lu, J. P., Si, Q., Kim, J. H. & Levin, K. *Phys. Rev. Lett.* **65**, 2466–2469 (1990).
19. Littlewood, P. B., Zaanen, J., Aeppli, G. & Monien, H. *Phys. Rev. B* **48**, 487–498 (1993).
20. Cho, J. H., Chou, F. C. & Johnston, D. C. *Phys. Rev. Lett.* **70**, 222–225 (1993).
21. Maeno, Y., Kakehi, N., Kato, M. & Fujita, T. *Phys. Rev. B* **44**, 7753–7756 (1991).
22. Koike, Y., Kawaguchi, T., Watanabe, N., Noji, T. & Saito, Y. *Solid St. Commun.* **79**, 155–158 (1991).
23. Axe, J. D. et al. *Phys. Rev. Lett.* **62**, 2751–2754 (1989).
24. Büchner, B. et al. *J. Low Temp. Phys.* **95**, 285–291 (1994).
25. Luke, G. M. et al. *Physica C* **185–189**, 1175–1176 (1991).
26. Kumagai, K. et al. *Physica C* **185–189**, 913–914 (1991).
27. Sternlieb, B. J., Tranquada, J. M., Shirane, G., Sato, M. & Shamoto, S. *Phys. Rev. B* **50**, 12915–12919 (1994).

ACKNOWLEDGEMENTS. We acknowledge discussions with V. J. Emery, J. Zaanen and S. Kivelson, and valuable assistance from B. Nachumi, Y. J. Uemura and G. M. Luke. The work at Brookhaven was supported by the Division of Materials Sciences, US Department of Energy, while that at the University of Tokyo was supported by a Grant-in-Aid for Research from the Ministry of Education, Science, and Culture of Japan.

Carbon nanotubes as removable templates for metal oxide nanocomposites and nanostructures

P. M. Ajayan*, O. Stephan*, Ph. Redlich**† & C. Colliex*

* Laboratoire de Physique des Solides associé en CNRS, Université Paris-Sud, Bâtiment 510, 91405 Orsay, France

† Max-Planck-Institut für Metallforschung, Institut für Werkstoffwissenschaft, Seestrasse 92, D-70174 Stuttgart, Germany

SEVERAL techniques have been developed recently for fabricating nanocomposite structures by filling^{1–3} carbon nanotubes^{4,5}. Here we show that surface-tension effects can induce the growth of uniform, thin metal oxide films—sometimes only a monolayer thick—on the outside of nanotubes, along with oxide fillings in the internal cavities and thin oxide layers between the concentric shells of the tubes. We report the preparation of such nanotube–oxide composites by annealing a mixture of partially oxidized nanotubes and V_2O_5 powder in air above the melting point of the oxide. The external coatings of the tubes are crystalline sheets of the V_2O_5 layer-like structure, which grow with the c axis parallel to that of the nanotube layers. Intercalation of the oxide occurs where there are missing shells in the nanotubes. We also show that the nanotubes can be partially removed by oxidation, leaving behind layered oxide fibres. Given the importance of vanadium oxides as catalysts and functional ceramics, this role of nanotubes as removable templates might lead to useful new kinds of nanostructured materials.

Carbon nanotubes produced by the arc-discharge method⁵ were ground and partially oxidized in air⁶ or nitric acid² to open some of the nanotube tips. The oxidized sample was then mixed and ground with pure V_2O_5 powder (dark orange) in approximately 1:1 weight ratio. The mixture was heated in an oven in air at $\sim 750^\circ\text{C}$ (melting point of the oxide is 690°C) for ~ 20 minutes. The annealed mixture (black) was crushed, dispersed in ethanol and observed by transmission electron microscopy (TEM) (Topcon 0020B) and scanning transmission electron microscopy (STEM) (VG HB 501).

We found that a large fraction of the open nanotubes were filled with the oxide. The fillings have diameters as small as 1–2 nm and extend continuously over several hundred nanometres. In cavities smaller than ~ 3 nm, the filling has an amorphous

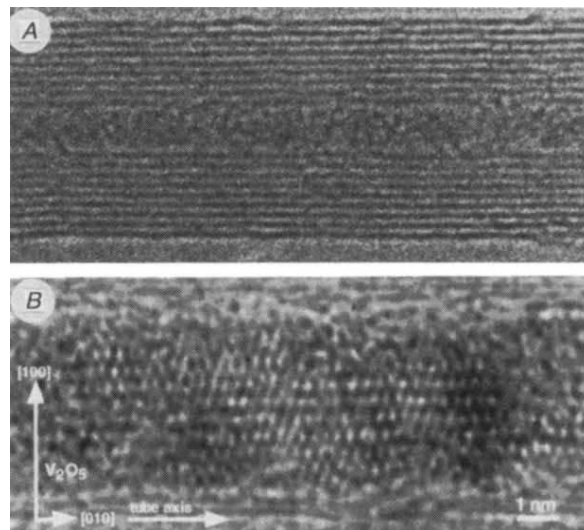


FIG. 1 High-resolution TEM images of tubes filled (darker contrast in the centre) with vanadium oxide. A, Filling in a narrow internal hollow, showing amorphous structure. B, Crystalline V_2O_5 structure in a larger hollow (only 2–3 tube layers around the filling is shown) showing preferred orientation with respect to the tube axis.

structure and the contrast in the images suggests that it has high porosity (Fig. 1A). In larger cavities however, the filled phase corresponds to crystalline V_2O_5 , and in some cases the structure clearly exhibits a preferred orientational relationship with the tube axis (Fig. 1B; the b axis of V_2O_5 is parallel to the tube axis), suggesting directional solidification of the molten oxide in the cavity. This striking change in solidification behaviour as the cavity size is reduced (stabilization of amorphous phase; also see ref. 1) is in agreement with recent theoretical predictions⁷ of filling with lead. From capillarity studies of nanotubes⁸, a surface tension of less than $100\text{--}200\text{ mN m}^{-1}$ is needed for any material to wet (and fill) nanotube surfaces. Molten V_2O_5 has extremely low surface tension ($\sim 80\text{ mN m}^{-1}$, comparable to water which has a value of $\sim 72\text{ mN m}^{-1}$)⁹ and should strongly wet nanotubes. In fact, the only pure binary oxides that satisfy the above wetting criteria are PbO , Bi_2O_3 , V_2O_5 , WO_3 , MoO_3 and B_2O_3 ; the first three have all been reported to fill nanotubes by capillarity (refs 1, 6 and this work).

As well as observing capillarity-induced filling of nanotubes, we find many isolated nanotubes covered (also by capillarity)

FIG. 2 Images of oxide-filling/nanotube/oxide-coating sandwich structures. a, Lower-magnification image showing uniform coatings on the left (completely covered for more than a few hundred nanometres) and right parts of the tube, with no coating in the centre. Filling is also seen inside the tube. b, Image at higher magnification resolving the 0.34-nm (002) fringes of the carbon nanotube. Coatings of oxide, uniform in thickness, can be seen on the upper and lower parts of the tube. The cylindrical profile of the coating is seen on the right side of the image. A diagram of capillarity-induced filling and coating of open tubes is shown at the bottom of the figure.

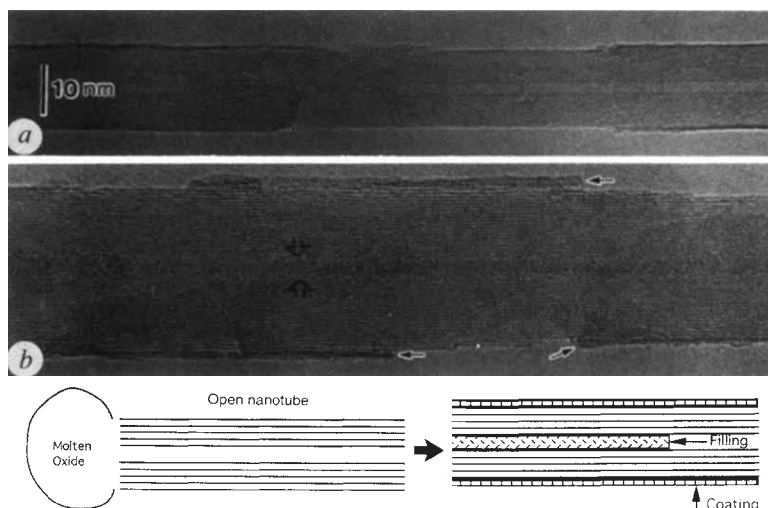
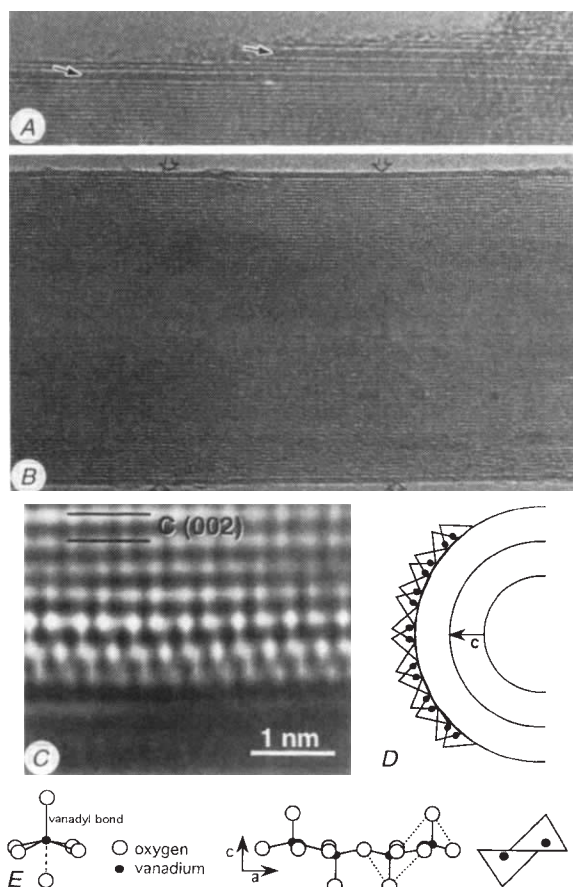


FIG. 3 High-resolution TEM images of nanotubes after heating with V_2O_5 at 750 °C. A, Intercalation by V_2O_5 in the gaps where nanotube layers are missing (arrows). Only a few outer fringes of a larger tube are shown. B, Cross fringes in the nanotube image, arising from the interference between (002) planes of graphite (0.34 nm) and V_2O_5 lattice in [001] projection. C, Higher-magnification image of the fringe structure due to the oxide coating at the edge of the tube shown in B (arrows). The image is Fourier filtered. D, Diagram showing the arrangement of structural pyramidal units on the surface of the tube, viewed parallel to the tube axis. The outermost circle drawn does not represent a nanotube shell. The arrow indicates the direction of the *c* axis in the nanotube. In this case, the V_2O_5 layer is projected in the [110] direction. E, Diagram of a basic pyramidal unit of V_2O_5 and a projection of a monolayer onto the *b* plane, showing upward and downward pointing pyramids. In D triangles representing pyramids point only outwards, because we assume that the inner vanadyl oxygens are missing. F, Analysis of the projected image of a coated tube acquired using a slow-scan CCD camera. The Bragg filtered diffractogram (inset) shows the superposition of the (002) spots (labelled 'C') from the carbon tube and the (110) and (200) spots corresponding to the overlayer V_2O_5 . The image is reconstructed in G by use of these spots. The contrast due to the oxide is clearly visible in the central hollow region. It matches well with a simulated image (inserted in G) of one monolayer of V_2O_5 in [001] projection at the optimum defocus (Scherzer) condition for phase-contrast imaging.



with extremely thin films of the oxide. In the sample, we also observe nanotube bundles covered with thicker deposits of the oxide. Figure 2 shows 'sandwich' structures of individual nanotubes which have inner fillings and thin outer oxide coverings. The coatings on isolated nanotubes are uniform in thickness and are rarely observed to be thicker than 1 nm. The coatings do not always cover the entire nanotube surface but continuous cylindrical films are seen to extend for lengths up to a few hundred nanometres. Commonly, films are not observed on the surfaces of very small tubes (<5 nm) nor on the tips of nanotubes and nanoparticles.

We often observe intercalation of V_2O_5 in the gaps (as small as twice the tube layer spacing) left by partially missing shells (owing to scroll-like defects¹⁰ or removed during oxidation) (Fig. 3A). No inter-tube-layer expansion is seen, as the adjacent cylindrical layers are fixed. Figure 3B shows a typical high-resolution TEM image from a larger tube coated with monolayer film. The image shows cross fringes arising from the interference between the pseudolattice from the carbon nanotube (002) layers and the overlying oxide lattice. The outermost fringe in the image corresponds to the oxide film coating (Fig. 3C) and represents the projection from the curved V_2O_5 layer (built from VO_5 pyramidal structural units sharing edges and corners (Fig. 3E))^{11,12} decorating the tube surface (see Fig. 3D). V_2O_5 has a layer-like structure, with the layers held together by weak V–O (vanadyl, 0.28-nm) bonds. We suggest that in the monolayer film only outward-pointing pyramids exist, leading to vanadyl oxygen vacancies, which has been observed for V_2O_5 surfaces^{11–14}. The missing vanadyl oxygens (discussed later) would give greater freedom for the layers to bend¹¹ and would allow a better match between the pyramidal bases of the V_2O_5 and the carbon lattice (see the epitaxial relation described later). For the intercalated films, this leads to the possibility of flat basal planes with no vanadyl oxygens sticking out of the

plane. Because of the tilt of the tube with respect to the imaging direction¹⁵ or actual bending of planes in the film, we often observe fringes due to the film curving around the tube axis.

To understand the symmetry and orientation of the coated film on nanotubes and to minimize radiation damage in the oxide layer, images from tubes with monolayer coatings (Fig. 3F) were acquired using a slow-scan CCD (charge-coupled device) camera. The Fourier transform of the image (inset) shows (002) spots of the nanotube along with spots corresponding to (200) and (110) planes of V_2O_5 , which has an orthorhombic unit cell¹¹, in the [001] (*c*-axis) projection. The relative interplanar distances and angles are correct within an error of 5% which could be due to distortions in the curved lattice. The contrast seen in the area corresponding to the central nanotube hollow of the Bragg filtered image (Fig. 3G) compares well with a simulated image (inserted in Fig. 3G) of one layer of V_2O_5 . The missing vanadyl oxygens in the structure do not alter the

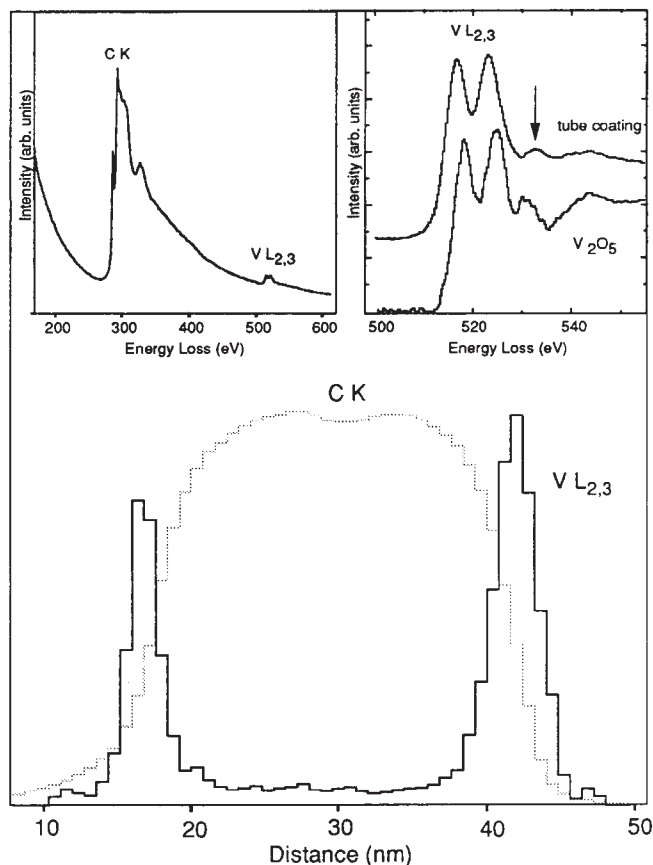


FIG. 4 Electron energy-loss spectra from coated nanotubes. Core loss spectra showing graphite (K) edge due to the nanotube and the vanadium (L) edge due to the coating (left inset). The oxygen K-edge at 532 eV is superimposed on the V L-edge. The fine structure (right inset) of the oxide L-edge suggests oxygen deficiency in the V_2O_5 coated films (for comparison, fine structure from bulk V_2O_5 taken using the same scan interval (0.2–0.8 s) is shown). The energy width between the L-edges and the oxygen prepeak (arrow) is larger for the coated film (typical signature for oxygen loss) and the $L_{2,3}$ peaks look symmetrical compared to that from pure V_2O_5 . The main figure shows intensity profiles calculated from line spectra across a 25-nm tube for the V L-edge and C K-edge. The profiles indicate the presence of a thin layer of vanadium around the surface of a large nanotube with a small hollow.

image contrast in this projection. The result suggests that during melting, wetting and solidification of the oxide on nanotubes, individual sheets have topotactically grown on the tube surfaces with their c axis normal to the curved graphite layers.

The layered nature of the oxide seems to help in obtaining such fine and uniform coatings because individual sheets slide easily with respect to each other. During intercalation of V_2O_5 , the layers can buckle¹² where the pyramidal units share only corners, suggesting that the sheets could effectively bend to cover the nanotubes. On the smallest tube on which coating was observed (~ 5 nm), this corresponds to 8.2° bending between pyramids, and 44 such pyramids are needed to cover one tube circumference if the film has its a axis parallel to the tube axis. After analysing many images of tubes with coated films, we fail to see any preferential orientation of the a and b axes of the V_2O_5 sheets with respect to the tube axis, although there could be orientational alignment between the film and the substrate lattices. It is impossible at this stage to assess this relationship because the helicity of sheets of carbon atoms⁴ in the outer layer of a multilayer tube cannot be uniquely determined.

The existence of oxygen vacancies in the film, such as missing vanadyl oxygens, is consistent with our results from electron

energy-loss spectroscopy (EELS) and observations^{13,14} on V_2O_5 indicating the formation of vanadyl surface defects after heat treatment. Spatially resolved line scans collecting electron energy-loss spectra¹⁶ across coated tubes (Fig. 4) gives concentration profiles indicating surface layers of vanadium oxide around carbon tubes. EELS from coated and filled tubes give signals corresponding to graphite and vanadium oxide¹⁷. For the vanadium L-edge, fine structure (Fig. 4 inset) is characteristic of oxygen-deficient stoichiometry compared to pure V_2O_5 . However, the reduced dimensionality of the oxide and radiation-induced oxygen loss^{18,19} may also contribute; but to minimize the latter we use very small scanning time intervals (see Fig. 4 legend).

As carbon nanotubes oxidize below the melting point of V_2O_5 (ref. 6), we attempted to see if the nanotubes can be preferentially removed by oxidation to leave behind some of the coated films and fillings. A second oxidation run at 650 – 675°C in air was carried out on the annealed sample for ~ 45 minutes, until the sample turned orange-grey (close to V_2O_5 colour) and the

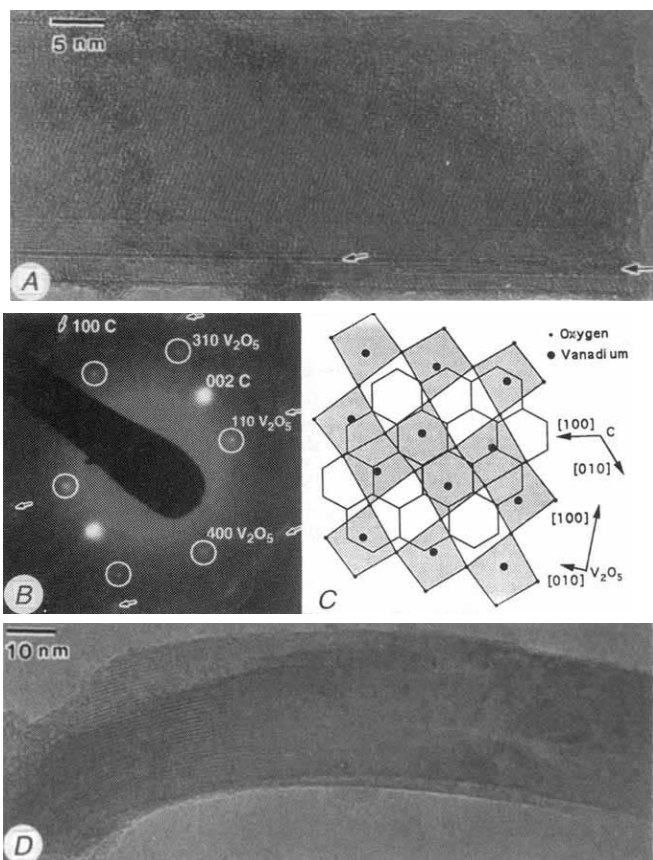


FIG. 5 Images from twice-oxidized sample (see text) showing the selective removal of carbon nanotube templates. A, Tip of a coated tube showing etching of the nanotube layers from the inside out. The coated film is seen near the open tube end (top right in the image) where the carbon tube has burnt away. Typical fringe contrast due to the outer oxide film is clearly visible because of the expanded tube hollow. The initial size of the hollow in this tube was only ~ 4 nm (seen away from the tip, but not shown in this image). Intercalation by the oxide is also seen in this tube (indicated by arrows). B, Electron diffraction pattern taken from the upper part of tip region of this tube (digitized and contrast-enhanced pattern). Faint spots due to V_2O_5 are marked with circles. Only the inner circle of $(hk0)$ spots is shown (indicated by arrows). C, Superposition of the distorted square net of the oxide lattice and the graphite hexagons, corresponding to the electron diffraction pattern in B. D, Image of a fibrillar structure of pure V_2O_5 left behind after the supporting nanotube has been removed. The structure is partially bent because of electron irradiation.

sample quantity reduced by about half. Among the few nanotubes left behind in this twice-oxidized sample, the ratio of coated to uncoated tubes had increased dramatically (>80%). Observation of the tips of the remaining coated tubes showed that, in most cases, oxidation has preferentially etched the inner layers of the tube, leaving the outer coatings and few outer tube layers (Fig. 5A) and giving clearer views of the coated film through the enlarged tube centres. This contrasts with the oxidation behaviour of pure nanotubes, where the attack proceeds from the outside to the inside, leaving chisel-shaped tips.

A possible orientational relationship between the outermost tube-layer lattice and the coated film now becomes discernible in some cases (using electron diffraction; Fig. 5B), owing to the selective removal of most of the inner layers. The few remaining outer tube layers (Fig. 5A) produce a narrow angular spread for the (*hk*0) spots ($\sim 4^\circ$), and the approximate orientation of the outer carbon honeycomb lattice can be determined with respect to the tube axis. The orientation of the oxide lattice can be deduced from the visible V_2O_5 spots in the diffraction pattern. In this case the (001) V_2O_5 basal planes grow on the (001) graphite surface with a near-parallel coincidence for two sets of planes; the (100) graphite|| (110) V_2O_5 , and (010) graphite|| (310) V_2O_5 (Fig. 5B, C). For the second set of planes the spacing mismatch ($\sim 20\%$) is in the range that could introduce van der Waals epitaxy^{20,21} in the system, as observed in layered dichalcogenides and Langmuir–Blodgett films. This type of lattice alignment occurs between weakly interacting surfaces with no dangling bonds (on cleavage planes, as here).

After the second oxidation, we find that the sample contains mainly larger (micrometre-sized) oxide structures, which covered nanotube bundles initially but now show pure oxide composition and hollow morphology (from EELS mapping), all graphitic structures having been removed. There are many extremely thin oxide skins (typically 10–50 nm in width, up to few hundred nanometres long and <1 nm thick) left behind in the sample. We see a few cylindrical structures of dimensions typical of larger nanotubes (Fig. 5D) of pure V_2O_5 (verified by diffraction and EELS). The observed skins could have originated from discontinuous films on nanotubes or from continuous films broken during sonication. The nanofibrils of V_2O_5 are flexible and collapsible under electron irradiation. We also find some solid rod-like V_2O_5 structures with a morphology typical of the filling in tubes, typically 5–8 nm in diameter and a few hundred nanometres long. The results indicate that nanotube templates are partially removable and point to their possible use as degradable templates to fabricate new ceramic films having layered structures.

The ability to coat the unreactive graphite basal planes that make nanotube surfaces with thin V_2O_5 films increases the versatility of nanotubes, because V_2O_5 is a good building block for catalytic complexes by organic routes²². Coated conventional carbon fibres have found a variety of applications²³. Coated and filled nanotube sandwich structures should have vastly different optical and electrical properties, and have better oxidation resistance (from our observations) than pure nanotubes. Vanadium pentoxide is an important catalyst²⁴ and a functional ceramic with uses ranging from batteries to switching elements^{25,26} with properties depending closely on dimensionality and oxygen stoichiometry. There is a great industrial interest in growing thin films of different vanadium oxides²⁷. Our results point to a new synthesis route for the fabrication of such thin-film ceramic nanostructures using nanotubes as efficient templates. □

Received 20 April; accepted 22 May 1995.

1. Ajayan, P. M. & Iijima, S. *Nature* **361**, 333–334 (1993).
2. Tsang, S. C., Chen, Y. K., Harris, P. J. F. & Green, M. L. H. *Nature* **372**, 159–162 (1994).
3. Guerret-Piécourt, C., Le Bouar, Y., Loiseau, A. & Pascard, H. *Nature* **372**, 761–765 (1994).
4. Iijima, S. *Nature* **354**, 56–58 (1991).
5. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
6. Ajayan, P. M. et al. *Nature* **362**, 522–525 (1993).
7. Prasad, R. & Lele, S. *Phil. Mag. Lett.* **70**, 357–361 (1994).

8. Dujardin, E., Ebbesen, T. W., Hiura, H. & Tanigaki, K. *Science* **265**, 1850–1852 (1994).
9. Ikemiyu, N., Umemoto, J., Hara, S. & Ogino, K. *ISIJ Int.* **33**, 156–165 (1993).
10. Amelinckx, S. et al. *Science* **267**, 1334–1338 (1995).
11. Galy, J. J. *Solid St. Chem.* **100**, 229–245 (1992).
12. Cava, R. J. et al. *J. Solid St. Chem.* **65**, 63–71 (1986).
13. Oshio, T., Sakai, Y. & Ehara, S. *J. Vac. Sci. Technol.* **B12**, 2055–2059 (1994).
14. Smith, R. L., Lu, W. & Rohrer, S. *Surf. Sci.* **322**, 293–300 (1995).
15. Cullen, S. L., Boothroyd, C. B. & Humphreys, C. J. *Ultramicroscopy* **56**, 127–134 (1994).
16. Ajayan, P. M. et al. *Phys. Rev. Lett.* **72**, 1722–1725 (1994).
17. Lin, X. W. et al. *Phys. Rev. B* **47**, 3477–3481 (1993).
18. Smith, D. J., McCartney, M. R. & Bursill, L. A. *Ultramicroscopy* **23**, 299–304 (1987).
19. Fan, H. J. & Marks, L. D. *Ultramicroscopy* **31**, 357–364 (1989).
20. Zasadzinski, J. A. et al. *Science* **263**, 1726–1733 (1994).
21. Ohuchi, F. S., Parkinson, B. A., Ueno, K. & Koma, A. *J. appl. Phys.* **68**, 2168–2175 (1990).
22. Gai, P. L. & Kourtakis, K. *Science* **267**, 661–664 (1994).
23. Dresselhaus, M. S. et al. *Graphite Fibers and Filaments* (ed. Cardona, M.) 238–243 (Springer, Berlin, 1988).
24. Krylov, O. V. *Catalysis by Nonmetals* (ed. Loebel, E. M.) (Academic, New York, 1970).
25. Scarmiento, J. et al. *Electrochim. Acta* **38**, 1637–1642 (1993).
26. Kim, D. H. & Kwok, H. S. *Appl. Phys. Lett.* **65**, 3189–3191 (1994).
27. Shimizu, Y., Nagase, K., Miura, N. & Yamazoe, N. *Jap. J. appl. Phys.* **29**, L1708–L1711 (1990).

ACKNOWLEDGEMENTS. We thank P. Ballongue and C. Jeanguillaume for their help. P.R. acknowledges a grant from the DAAD (HSP-II).

Deflation of Mount Etna monitored by spaceborne radar interferometry

Didier Massonnet*, Pierre Briole† & Alain Arnaud‡

* Centre National d'Etudes Spatiales, 18 Avenue E. Belin, 31055 Toulouse, France

† Institut de Physique du Globe, 4 Place Jussieu, 75005 Paris, France

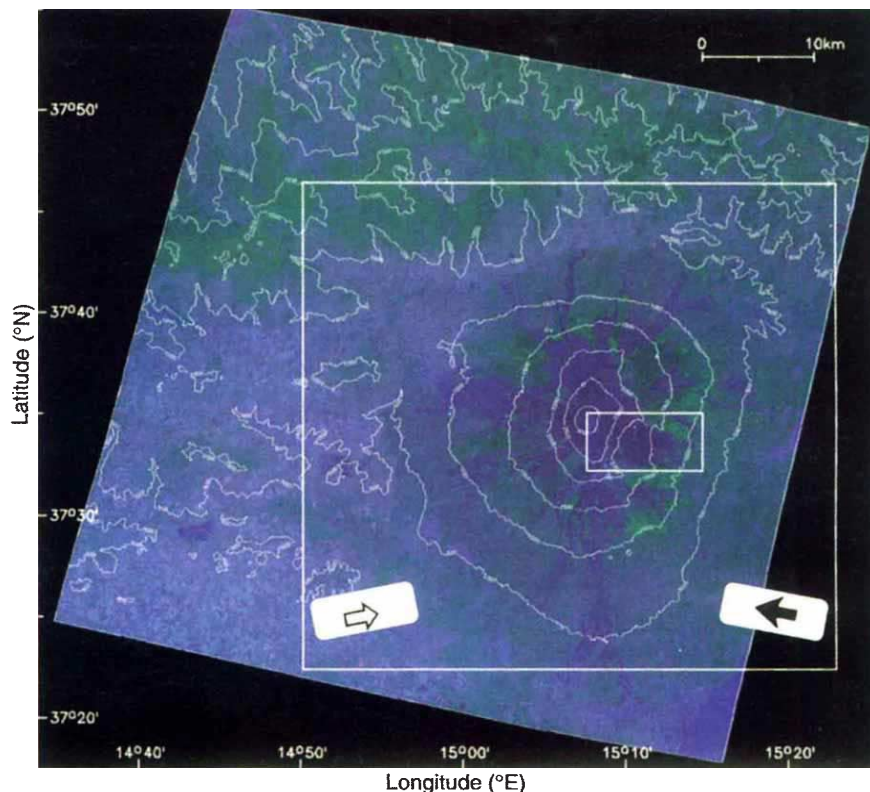
‡ Groupe CISI, 1 Rue des Cosmonautes, 31400 Toulouse, France

GROUND-BASED measurements of volcano deformation can be used to assess eruptive hazard, but require the costly (and often hazardous) installation and maintenance of an instrument network. Here we show that spaceborne radar interferometry, which has already shown its utility in mapping earthquake-related deformation¹, can be used to monitor long-term volcano deformation. Two families of synthetic aperture radar images, acquired from ascending and descending orbits by the satellite ERS-1, and looking at Mount Etna from opposite sides, cover the time period from 17 May 1992 to 24 October 1993, and include the second half of Etna's most recent eruption. Despite artefacts of the interferometric technique, we can observe a volcano-wide deflation, which is an expected consequence of the eruption, but which had not previously been appreciated. We quantify it using a simple model based on the change of pressure in a sphere located in an elastic half-space; the modelled deformation increases linearly with time until the end of the eruption. Our results show that it will be possible to use this technique to detect the inflation of volcanic edifices that usually precedes eruptions.

Mount Etna (Fig. 1) is one of the most active and best studied volcanoes in the world. The last eruption started on 14 December 1991 in the Valle del Bove, a large amphitheatre formed by collapse of the eastern flank. Lava was erupted along a fracture system that had opened in 1989², and covered most of the southern part of the Valle del Bove. The eruption stopped on 31 March 1993 after 473 days. The rate of lava production remained stable during most of the eruption, and the total erupted volume was $\sim 3 \times 10^8 \text{ m}^3$ (ref. 3).

To measure the deformation, we constructed interference patterns from the difference of pairs of synthetic aperture radar (SAR) images acquired by the ERS-1 satellite. This geodetic technique can measure centimetre-size changes in the ground surface^{1,4,6}. The resulting interferogram is a contour map of the change in range to the volcano surface, measured along the

FIG. 1 Location map from a SPOT image, representing the limits of the digital elevation model used with ascending as well as descending radar images. The model has been computed by ISTAR S. A. (Valbonne, France) using two SPOT images acquired on 22 and 23 July 1991. Vegetation-covered areas appear in green and lava flows are dark. The radar antenna points roughly to the West in ascending orbits and to the East in descending orbits. Interferometric sensitivity to various ground displacements can be determined from unitary vectors pointing from Mount Etna towards the satellite in its ascending or descending orbits. These vectors, the Earth-surface projections of which are shown (outlined arrow, ascending; solid arrow, descending), are given explicitly as (0.418, -0.085, 0.904) for ascending orbits and (-0.373, -0.077, 0.925) for descending orbits in a coordinate set (east, north, up) on the Earth's surface. These satellite-directed vectors indicate roughly the same sensitivity of the two data sets to south-north or vertical displacements, but reveal opposite signs for east-west displacements. The values given above for the summit of Mount Etna. Their variations within the image frame and within a given family of orbits are very slight and can be neglected. The larger box outlined in white is the location of Figs 3 and 4. The area where local surface deformation can be seen in Fig. 3a, e, h and 4a is indicated by the smallest box.



direction towards the satellite. Each contour line, or fringe, represents 28 mm, or half the wavelength of the radar. Our approach requires two images and uses a digital elevation model to calculate and remove the effect of topography. For this study we used a digital elevation model of the area accurate to 8 m (r.m.s. error), according to its technical documentation (ISTAR, Valbonne, France). The model has been obtained using data from the optical remote-sensing satellite SPOT. The accuracy of the measurement is limited mostly by atmospheric propagation heterogeneity and residual topographic contributions. Propagation heterogeneity introduces errors that are generally less than

half a fringe, but which may be as large as three fringes locally, in exceptional conditions. A pair-wise logic allows their identification as tropospheric or ionospheric artefacts⁷. Residual topographic errors depend on the separation of the two orbits: the smaller the separation, the lower the sensitivity to topography and any residual topographic errors. An interferogram is characterized by the time interval it spans and by its altitude of ambiguity⁸, equal to the amount of elevation change that produces one topographic fringe. The altitude of ambiguity is roughly inversely proportional to the orbital separation.

The ERS-1 satellite acquired 13 radar images of Mount Etna during ascending orbits (travelling north), looking roughly to the east during the night (23:16 local time). It also acquired 16 radar images during descending orbits (travelling south), looking roughly to the west during the day (11:40). Only images from the same family of orbits can be combined. The ascending (descending) images give 78 (120) potential interferometric combinations. We selected 32 ascending and 60 descending combinations with altitudes of ambiguity larger than 20 m, thus ensuring negligible residual topographic contribution. About 30 interferograms presented enough coherence to be analysed and 12 of them contributed quantitatively to the present study. Their features are summarized in Fig. 2.

Experience with this large interferometric data set has led to three general conclusions: on average the interferograms made from ascending (night-time) passes are better than those from descending passes, possibly owing to more stable atmospheric conditions and a more quiescent vegetation state; interferogram phase changes due to ionospheric or tropospheric propagation effects⁷ are common and must be identified to avoid displacement errors of up to one fringe, or 28 mm (Fig. 3d, e, h, i); and the coherence between images does not decrease systematically with time. Fig. 4a shows one of the best ascending interferograms, obtained over 385 days. The image is dominated by four concentric fringes that suggest an ~11-cm deflation of the whole volcano. Localized fringes are also visible in the Valle del Bove and show local subsidence, well correlated with the 1989 lava

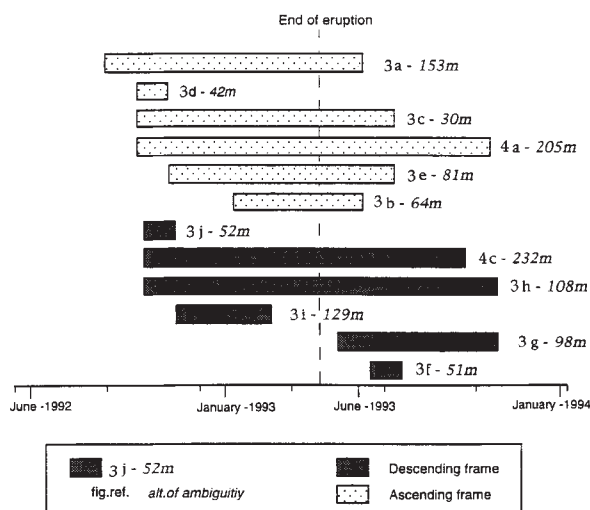


FIG. 2 Bar chart of the interferometric pairs shown as figures in this study. Ascending pairs are represented by dotted bars, descending pairs by shaded bars. The numbers on the right-hand side of each bar are the figure code of the interferograms and the altitude of ambiguity.

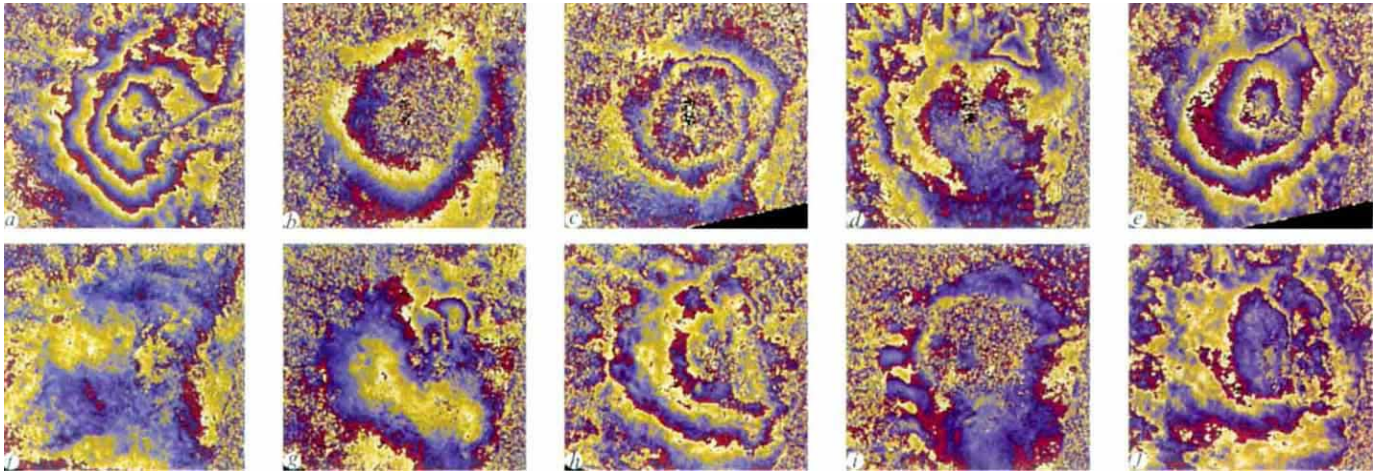


FIG. 3 Examples of interferograms (a–e, ascending orbits; f–j, descending orbits). a–c Show the subsidence of the volcano at different stages. The triangular feature in d and e is related to the ascending orbit of 1 December 1992, and is equivalent to a positive displacement of 3 cm. It does not appear in the interferogram 3c and is thus produced by a reversible source. This feature also does not appear in interferogram j that is made from images taken on 4 October and 8 November 1992. The most likely cause is thus an ionospheric propagation

heterogeneity⁷. Interferogram f (35 days separation, starting three months after the end of the eruption) shows good coherence and not a single fringe despite a relatively small altitude of ambiguity (51 m). It was used to assess the quality of the digital elevation model and thus to calibrate the maximum expected topographic residual on the other interferograms. Examples of tropospheric heterogeneities are shown in h (semicircular feature different from a closed fringe) and i (chained 'clouds' on the left-hand side).

field. Such phenomena have been observed on Etna by Murray⁹ near recent lava flows, and can be interpreted as compaction of the fresh unconsolidated lava flow and downward flexure around it. The same subsidence is visible on the interferograms shown in Fig. 3a and e.

Deflation is not observed on the interferograms formed by images acquired after the end of the eruption, such as Fig. 3f. This interferogram is uniform in the areas where coherence is kept. Because its altitude of ambiguity is 51 m, we expect at

most one-sixth of a cycle of topographic residual from the 8 m credited accuracy of our elevation model, which is clearly the case here. Such a topographic residual would account for less than 1/25 of a fringe, or 1 mm, in an interferogram with an altitude of ambiguity of 205 m, as Fig. 4a. Similarly, the fringes in Fig. 4a cannot be due to atmospheric propagation heterogeneities attached to one image, as the almost synchronous interferogram Fig. 4c shows a deflation of similar amplitude although formed with two images from descending orbits, which could

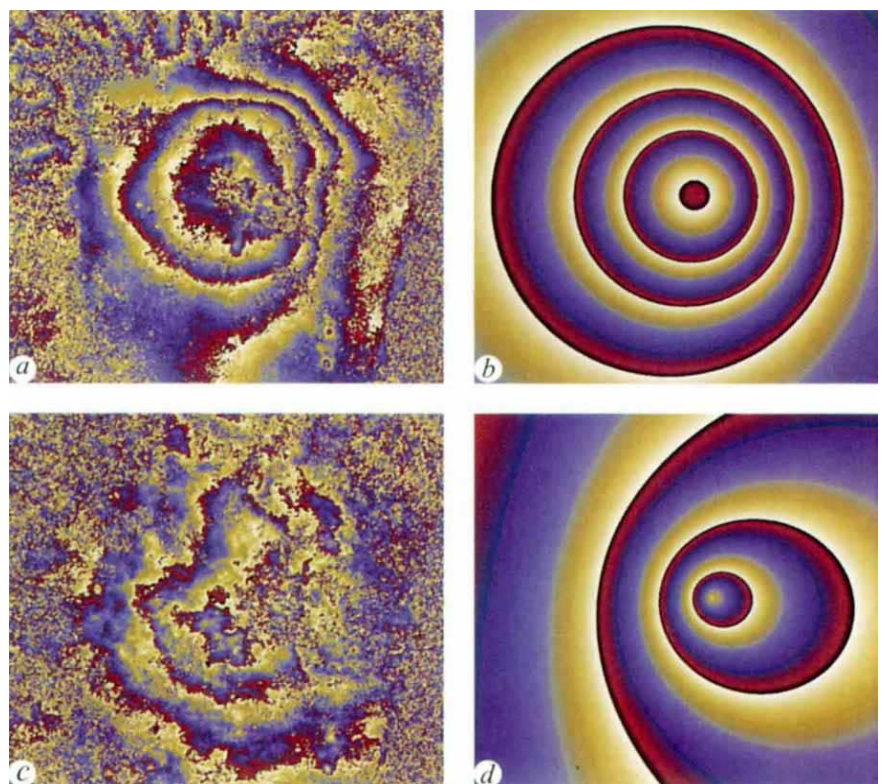


FIG. 4 Actual and modelled fringe patterns for the best ascending and the corresponding descending interferograms. One cycle of colour corresponds to a 28-mm change in range. About four fringes are visible in a, indicating a subsidence of the volcano. The interferogram corresponding to the best-fit model is shown in b. An almost synchronous descending interferogram is shown in c, and d is the interferogram for the best-fit model. The best fit is obtained with a source located 2 ± 0.5 km east of the centre cone and at 16 km depth.

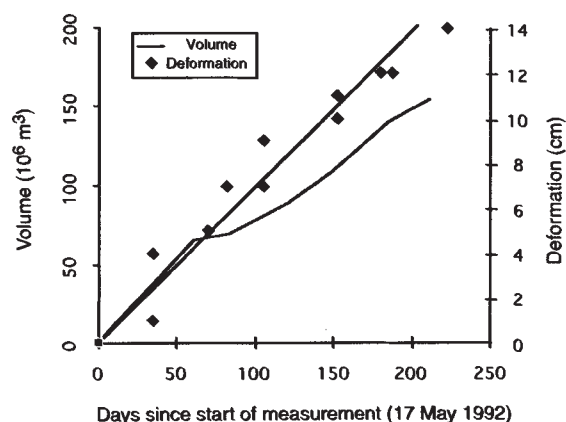


FIG. 5 Subsidence of the summit of Etna deduced from our data and plotted (diamonds and associated straight-line fit) as a function of the effective elapsed time. Volume of lava produced during the eruption (other solid line) as inferred from the field measurement³. Because no deformation is visible after the end of the eruption (Fig. 3f), the effective elapsed time was defined as the number of days that separate the first day of the interferogram and the last day of the interferogram or the last day of the eruption if earlier (Fig. 4). The error on the individual measurements is ≤ 1 cm, based on the quality of the model fit, assessed at intervals of 1 cm. Interferograms that did not provide an unambiguous fit at the 1-cm confidence level were not included in the quantitative study.

not have been affected by the same atmospheric propagation heterogeneities.

Assessing the deformation just by counting the fringes is difficult for several reasons. First, an interferogram does not give an absolute indication of phase and the background phase is difficult to assess. Second, the volcano-wide deflation pattern is obscured in areas where coherence is poor, such as some vegetation-covered areas and the summit, which freezes at times. The phase is also biased by local deformation and atmospheric artefacts. Thus, we separated the contribution of volcano-wide deflation by fitting the interferograms with a calculated theoretical displacement field derived from a simple model of pressure change in a sphere located in an elastic half-space¹⁰. The model, which also gives the total volume change, depends on four parameters: the coordinates of the centre of the sphere and the maximum amplitude of the displacement above this centre. In this model, we prescribe only changes of volume at depth, regardless of the consequences in terms of pressure, which would require knowledge of the size and rigidity of the source.

We used the ascending interferogram Fig. 4a, in combination with the best descending interferogram Fig. 4c, to fix the centre of the sphere. Its horizontal location was obtained by ascertaining the best fit of the centroids of the fringes for the data (Fig. 4a and c) to those for the models (Fig. 4b and d). The best fit was obtained with the centre of the sphere located 2 ± 0.5 km east of the central cone of Etna.

The maximum amplitude of subsidence in the interferogram shown in Fig. 4a was obtained by minimizing the residuals for a range of assumed depths of the sphere (10–20 km). The sensi-

tivity of the solution to the depth is low, and our best fit was always obtained with a subsidence of 12 cm. This value is well constrained, in particular by the western and the southern parts of the image. We estimate the uncertainty in this value to be of the order of 1 cm. For the depth of the sphere, the best fit is not sharply defined, and is obtained between 13 and 19 km. We thus estimate the depth of the source to be 16 ± 3 km. We emphasize that our model does not constrain the location of the magma reservoir(s) beneath Etna or its finite shape, except that the model is consistent with a source at mid-crustal levels—deeper than anticipated from previous interpretations^{11,12} based on measurements that did not sample the entire deformation field.

When modelling additional interferograms, we did not have to change the location of the centre of the sphere, but simply to adjust the maximum displacement, which indicates the geometric consistency of the model. For each interferogram we computed 20 residual images obtained by increasing the amplitude from 0 to 19 cm by steps of 1 cm and selected the minimum. We retained the 12 interferograms for which this choice was unambiguous at the 1-cm confidence level.

Figure 5 shows the maximum displacement (right-hand ordinate) plotted as a function of time, until the end of the eruption. These values are consistent whether obtained from ascending or descending interferograms, despite the very different sensitivity of the two families of data to east–west displacements, as explained in Fig. 1. This further confirms the geometrical consistency of the model. No deformation is measured after the end of the eruption, as shown by the interferogram in Fig. 3g. During the last seven months of the eruption sampled by our data, the maximum subsidence increases linearly at a rate of 21 mm per month. From GPS data, the subsidence of the volcano for the first five months of the eruption (inferred from the difference between data for July 1990 and June 1992) has been estimated¹¹ as 19 cm at the summit and 6–8 cm at points located 6–10 km from the top. Assuming that all the observed displacement occurred during the five months of eruption, the average rate of subsidence was 31 mm per month. The maximum horizontal strain rate predicted by our model is -0.9 p.p.m. per month for opposite points located at 11 km from the top, a value close to the -0.75 p.p.m. per month obtained earlier in the eruption¹¹.

The constant value of our subsidence rate is consistent with the stable rate of lava production measured in the field³ and plotted in Fig. 5. The rate of volume change at the surface of the volcano deduced from our observation is 3.4×10^7 m³ per month. This rate is not balanced by the production of lava (1.9×10^7 m³ per month).

This analysis of data from Mount Etna validates the use of radar interferometry for volcano monitoring. Volcano-wide surface deformation is revealed and the associated volume change can be measured. The simplicity of our model allowed us to use many interferograms, even those of low quality—a practice further justified by the consistency of the amplitudes of subsidence. This technique could be used for surveying other active volcanoes, using existing (ERS-1 and J-ERS1) or future (ERS-2 and Radarsat) radar systems. Continuing work on other volcanoes, such as an unpublished study that we have initiated of Merapi volcano (Indonesia), shows that it is possible to obtain coherent images even in equatorial areas, where rapid surface changes occur owing to lush vegetation. □

Received 21 December 1994; accepted 5 May 1995.

1. Massonnet, D. et al. *Nature* **364**, 138–142 (1993).
2. Villari, L. (ed.) *Acta vulcanol.* **4**, 1–177 (1994).
3. Global Volcanism Network Smithsonian Inst. Bull. **10**(17)–**10**(5) (1991–1993).
4. Goldstein, R. M., Engelhardt, H., Kamb, B. & Frolich, R. M. *Science* **262**, 1525–1530 (1993).
5. Zebker, H. A., Rosen, P., Goldstein, R. M., Gabriel, A. & Werner, C. L. *J. geophys. Res.* **99**, 19617–19634 (1994).
6. Feigl, K. L., Sergeant, A. & Jacq, D. *Geophys. Res. Lett.* (in the press).
7. Massonnet, D. & Feigl, K. L. *Geophys. Res. Lett.* (in the press).

8. Massonnet, D. & Rabauts, T. *IEEE Trans. Geosci. Rem. Sensing* **31**, 455–464 (1993).
9. Murray, J. B. *J. Volcanol. geotherm. Res.* **35**, 121–139 (1988).
10. Mogi, K. *Bull. Earthquake Res. Inst. Tokyo* **36**, 34–134 (1958).
11. Nunnari, G. & Puglisi, G. *Acta vulcanol.* **4**, 101–107 (1994).
12. Bonaccorso, A., Velardita, R. & Villari, L. *Acta vulcanol.* **4**, 87–96 (1994).

ACKNOWLEDGEMENTS. We thank L. Gray, P. Segall, G. Wedge and A. Bonaccorso for their helpful comments and suggestions. More than half of the ERS-1 scenes used in this study were provided free of charge by the European Space Agency. This study was partially funded by the Natural Risk Reduction Program of the European Community.

Seismic anisotropy as an indicator of mantle flow beneath the Himalayas and Tibet

A. Hirn*, M. Jiang†, M. Sapin*, J. Diaz‡, A. Nercissian*, Q. T. Lu†, J. C. Lépine*, D. N. Shi†, M. Sachpazi*, M. R. Pandey§, K. Ma† & J. Gallart‡

* Département de Sismologie associé au CNRS, Institut de Physique du Globe, 4 place Jussieu, 75252 Paris, France
† Institute of Geology and Mineral Deposits, Chinese Academy of Geological Sciences, Beijing, China
‡ Consejo Superior de Investigaciones Científicas, Barcelona, Spain
§ Department of Mines and Geology, Katmandu, Nepal

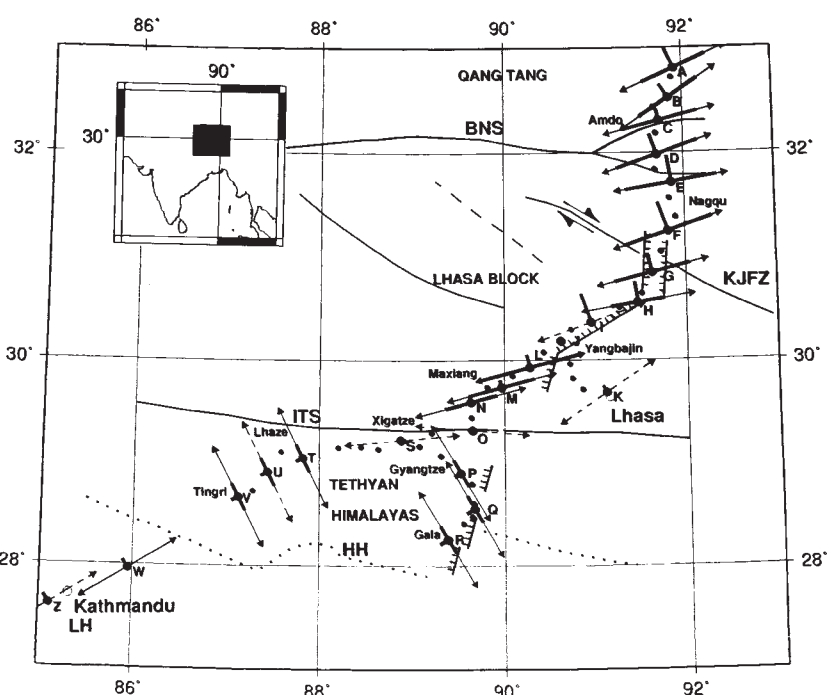
SEVERAL models have been proposed for the geodynamical evolution of the Tibet–Himalayas collision zone^{1–6}. It is now generally recognized that the high elevations of the region have been caused by mechanical thickening of the crust and flow in the mantle, but there is debate as to whether the thickening has occurred by the underthrusting of Indian crust under Tibet, or by distributed shortening and thickening of the Tibetan crust as India has pushed northwards into it. Here we address this question using seismic measurements of heterogeneity and anisotropy at depth, obtained with a temporary teleseismic array spanning 500 km from the Lesser Himalayas to central Tibet (Fig. 1). We observe a significant change in seismic anisotropy across the Indus–Tsangpo suture (ITS), suggesting a change in mode or direction of deformation at depth. In the Himalayas, our results are consistent with the stack-

ing of Indian and Tibetan lithospheres, whereas north of the ITS the data indicate ductile flow in the mantle and show no sign of the Indian lithosphere.

A seismic array with 60 stations was installed from the Lesser Himalayas in Nepal to Qang Tang in central Tibet in summer of 1992 (Fig. 1). A distributed subset of 25 sites had three-component Lennartz five-second (intermediate-period) sensors which recorded clear teleseismic shear waves on 250 event-to-station seismograms. Using this array, we have observed SKS shear-wave splitting, which is diagnostic of horizontal anisotropy at depth, and documented a strong variation of its characteristics along the array (Fig. 1). Teleseismic SKS waves travel through the Earth's outer core as compressional waves, then emerge into the mantle as shear waves. In a region of anisotropy (characterized by orthogonal 'fast' and 'slow' axes) such waves separate into two components, which propagate with different speeds. Standard methods have been developed^{7,8} to estimate, from recorded data, the direction of polarization of the fast wave and the splitting time (time-lag between arrivals at the recording station) of the fast and slow components (Fig. 2). Because SKS propagation is nearly vertical, observation of shear-wave splitting is diagnostic of azimuthal anisotropy beneath the recording station. SKS-splitting is thought to be due to a preferred orientation of olivine⁹ induced by the deformation of the lithospheric mantle⁷ or by flow in the asthenosphere^{8,10,11}, with the fast orientation giving the direction of flow. Other possibilities include the alignment of melt-filled cracks in a magmatic zone.

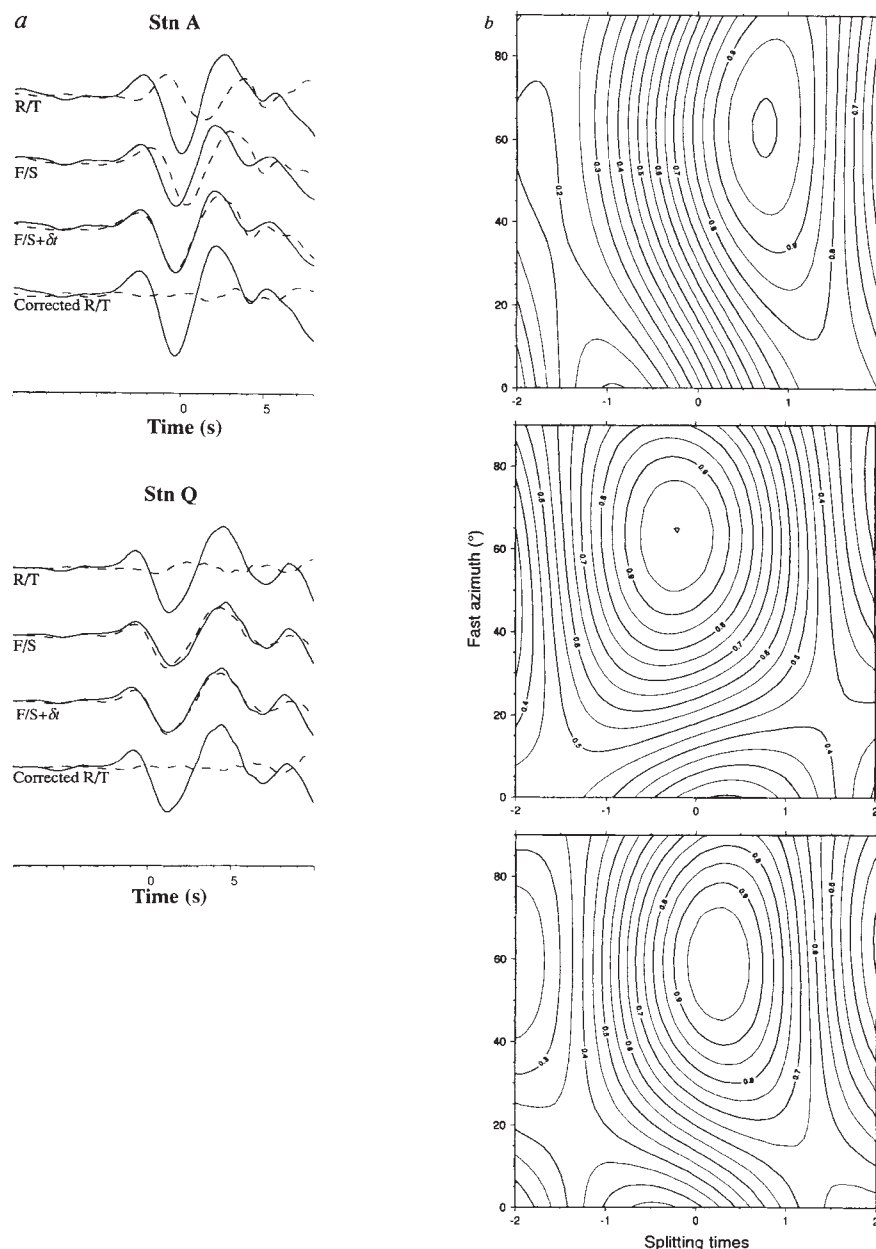
As shown in Fig. 1, the polarization of the fast wave is oriented towards the northeast in the Katmandu region, but changes across the High Himalayas (HH) to north-northwest in the Tethyan Himalayas. Its orientation changes again across the ITS to east-northeast in the Lhasa and Qang Tang blocks. No single

FIG. 1 Sketch of the Tibet–Himalayas collision zone showing sites of digital seismographs. Lettered sites are three-component, five-second sensors, others are one-component, one-second sensors. The teleseismic array was deployed from August to early November 1992 in Tibet and briefly in Nepal. Arrows at each site indicate fast directions of anisotropy (polarization of fast SKS wave) measured for a single earthquake (Vanuatu; distance 85°, back azimuth 110°) representative of the entire data base at the site. Dotted arrows at L, U and Z indicate that another earthquake was used because these sites were not operational for Vanuatu. At S and O most signals are not consistent with an isotropic or a tabular structure beneath the stations, showing deviation from theoretical polarization in the particle motion although ellipticity is not clear. At these stations, the observed polarization directions are plotted only with dots, but they may still indicate fast axes because they deviate from the 110° back azimuth expected in the absence of anisotropy. At K the criteria of splitting are not clearly met with the exception of one Vanuatu event which gives an obvious diagnostic of anisotropy as plotted here in dots. At each site, the length of the bold segment along the arrow represents the residual delay of the first SKS arrival relative to baseline station Z in Nepal. An absolute residual is obtained as the difference between the onset time of an observed P or S wave and that calculated from travel-time tables in a standard Earth for this distance from the earthquake. Subtracting the value at station Z gives a relative residual, and the relative residual shear delay plotted here is the relative residual for SKS minus 1.8 times the relative residual for PKP. The plotted delay is thus compensated for any structural variation with normal Poisson's ratio along the array, including the known change in topographic elevation, 1.5–5.5 km, and in crustal thickness from around 30 to 70 km. The observed spatial variations are consistent with variations of Poisson's ratio. Maximum relative residual delay is ~2.5 s at sites L and E. The length of the bold segment orthogonal to the arrow is the splitting time which is the time-lag between the slow and fast components of the SKS wave. The maximum is slightly <1 s at site F. Selected geological



features sketched along the array include the Jurassic suture zone of Banggong–Nujiang (BNS), between the Qang Tang and the Lhasa block, and the Eocene suture of the Indus–Tsangpo rivers (ITS), the Higher Himalayas (HH) (dotted line) and the Tethyan Himalayas, situated between HH and ITS. The Karakorum–Jiali (KJFZ) Quaternary shear zone of lateral escape of Central Tibet is indicated with strike-slip arrows; barbed lines are faults limiting grabens of Quaternary extension near the line of stations⁶.

FIG. 2 *a*, Orthogonal projections of the horizontal waveforms of the SKS wave from Vanuatu at stations A and Q. R (solid line) is the radial along the back azimuth; T (dotted line) is the transverse signal. The significant amplitude of the transverse arrival and its waveform, similar to the derivative of the radial arrival, are diagnostic of anisotropy. F (solid) and S (dotted) are the projections on the fast and slow principal directions of anisotropy obtained after a grid search (in *b*) for best fitting parameters. The fast direction is N65°E (full line) for A and N25°W (dotted line) for Q. F/S + δt shows a superposition of fast and slow components obtained by shifting by the splitting time δt . For A, $\delta t = 0.7$ s; for Q, $\delta t = 0.25$ s. Corrected R/T is the re-projection along radial (solid) and transverse (dotted) directions after having compensated for anisotropy effects. Note that taking anisotropy into account allows the recovery of the SKS wave without transverse component before it entered the anisotropic domain. *b*, Result of a grid search for anisotropy parameters (fast azimuth Φ and splitting time δt). Upper panel, station A in the north (fast axis 63° E, 0.7 s delay); middle panel, station Q south of the ITS (fast axis, 65°, -0.25 s delay); lower panel, station W in Nepal (fast axis, 60° E, 0.25 s delay). Contours represent constant values of the time-integrated, squared amplitude difference between the two orthogonal projections. Resolution can be estimated from isolines to be a few degrees in azimuth and about 0.1 s in delay.



mechanism of geodynamical evolution, supposed active throughout the region, can account for these observed changes in anisotropy. For instance, thickening of Tibet^{1,2} through underthrusting of the Indian crust should result in anisotropy orientation everywhere normal to the Himalayas¹² (across their strike), which is not observed north of the ITS. Coherent deformation of the crust and lithosphere⁷ (homogeneous thickening by vertical plane strain³) would be expected to give orientations along strike, which are not observed south of the ITS. The two stations situated less than 20 km south of the ITS have anomalous signals, polarized along the suture, caused probably by major lateral heterogeneity near the suture. For all other stations anisotropy is well defined, and the fast directions are consistent to better than 10° among stations in either region north and south of the ITS. Observed fast-wave polarizations that differ from those predicted by the simple geodynamical models discussed above may therefore be regarded as significant.

South of the ITS, in the Tethyan Himalayas, splitting is small. The across-strike orientation is N10°E, whereas observations show a consistent N25°W orientation (Fig. 1), which is the direction of relative motion between India and Tibet as inferred

from summing displacements on the diverse surface-tectonic elements in this region: the Himalayan frontal thrusts, the right-lateral Karakorum-Jiali fault zone (KJFZ) and the normal faults along several north-south grabens in southern Tibet^{6,13}. We presume that the orientation of fast polarization in this region reflects the deformation due to relative motion within the imbricated lithospheres that have given rise to the Tethyan Himalayas.

To the south of the High Himalayas only two stations are available. Their measurements indicate that, on the Indian plate, the magnitude of anisotropy remains small but has a fast orientation of N60° E. Although the current direction of motion of the Indian plate with respect to the hotspot reference frame is estimated to be due north¹⁴, this reflects an anticlockwise rotation following the first contact of India with Asia in the west. Before this collision, 45 Myr ago, India drifted towards Asia in a northeastward direction⁵. This could have corresponded to a large-scale mantle flow along which directional anisotropy could still be directed.

North of the ITS fast orientations are N70° E, as found previously at two sites within our array¹⁵. North of the Yangbajain

graben, our array has high resolution and crosses the east–west Banggong–Nujiang suture (BNS), which originated in the Jurassic period, and also the KJFZ, a quaternary, seismically active fault zone suggested to be a major lithospheric boundary⁶. The fast orientations are significantly different from the along-strike direction for each of these fault zones, and remain constant across them. To characterize anisotropy here we resort to additional data.

To provide a different insight into structural variation we may use near-vertical-incidence PKP compressional waves from earthquakes antipodal to the array to measure the variation of total crust and mantle transit times among sites (Fig. 3a). These results, which are explained roughly by variations in crustal thickness that we measured in an earlier wide-angle reflection experiment^{4,16}, imply that variability of mantle P-wave velocity is small. For shear-wave velocities, previous work has recognized regional variations but these were north of our survey region^{17–19}. For instance, the mean S-wave arrival-time residuals from the worldwide seismic network for major earthquakes show strong variations across Tibet²⁰. That study does not, however, resolve variation among the three events with epicentres spanned by our array.

Unlike the data for PKP waves, that for near-vertical-incidence SKS waves (Fig. 3b) reveals a strong increase in delay times north of the ITS, except to the east at station K in Lhasa. The cause of this increase of SKS delay in relation to PKP delay is unlikely to be the variation of the thickness of layers in the crust and mantle lithosphere, because in continental regions these layers have similar Poisson's ratios and corresponding ratios of P-wave to S-wave velocities. The rapid variation in the relative delay over a distance <100 km (north of the ITS or northwest of station K) suggests a shallow origin, probably in the crust and uppermost mantle. Thus a proper explanation appears to require a sharp lateral variation in Poisson's ratio. Geothermal fluids in the uppermost crust may contribute to this strong variation, but the most likely explanation is the presence of partial melt, as the relative reduction of S velocity as a function of melt fraction is larger than that of P velocity. If, to obtain a rough estimate, we supposed that the S-wave reduction is twice that of the P wave²¹, then the variation of two seconds of differential S–P residuals along the line can be accounted for by introducing 7% partial melt over an 80-km thickness, or other proportional combinations of these figures.

The region with large S delays and inferred partial melt reaches far south in the Lhasa block, evidence that cold cratonic India does not extend significantly north of the ITS, at least at shallow levels in the mantle. The slow region is bounded to the south of the Maxiang volcanics towards the ITS and also to the southeast of the Yangbajain graben, where the transition towards normal S delays at Lhasa is accompanied by a 70 mgal increase in the value of the Bouguer gravity anomaly²². The Maxiang suite of Miocene potassic calc-alkaline volcanism²³ is, along with the basalts of northern Qang Tang, the most recent in Tibet. It postdates collision and orogeny and has been proposed to represent magmatism related to lithospheric delamination²⁴. The Tibetan subcrustal earthquakes²⁵ have also occurred exclusively in this region south of Maxiang, and we infer that active tectonics and magmatism there occur at mantle depth. This zone is situated far south in Tibet, however, and also appears laterally more confined along strike than would be expected from proposed broad-scale delamination under central Tibet³.

The largest splitting times (one second) occur north of the ITS where the S wave arrives abnormally late. The presence of partial melt, which we suggest caused the observed S-wave delays, may also aid formation of anisotropy, as the associated heat would promote crystal reorientation under differential horizontal shear, or allow orientation of the fluid-filled defects by ductile flow. Nevertheless, the variations of S-wave delay and of splitting time are not completely correlated. From the ITS into

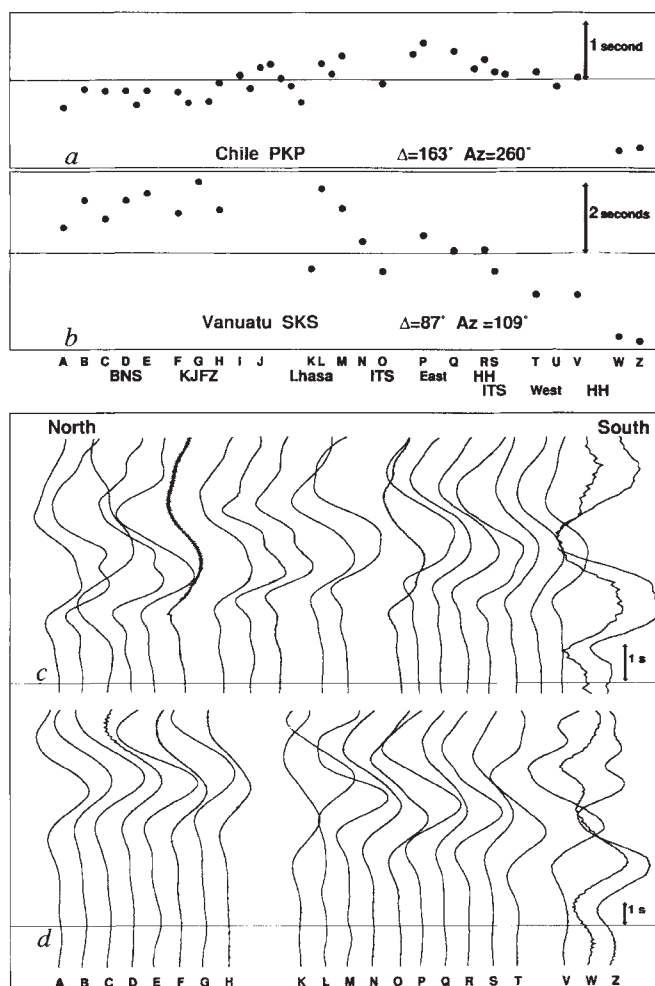


FIG. 3 Relative variation through the array of P or S travel-time residuals with respect to standard travel-time tables, corrected for different elevations. Note (from Fig. 1) that station K and the three one-component stations to the left of it are off the main line and that stations S to V are not aligned south of O to R along the main section, but form a second transect of the Tibetan Himalayas 150 km west of them to W, Z in Nepal. *a*, PKP waves at near-vertical incidence from an earthquake in Chile. These waves are very late in the Tethyan Himalayas between HH and ITS relative to those detected in Nepal in the South. Further north, crossing the ITS and KJFZ they become less late. This is consistent with the results of a wide-angle reflection traverse^{4,7} which showed a 75-km-thick crust between HH and ITS, double the thickness in Nepal, and suggested a 55-km thickness north of ITS averaged over significant local variations. *b*, SKS shear waves at near-vertical incidence from Vanuatu, on horizontal-component seismographs. The timescale is compressed by 1.8 so that the magnitude variation of relative residuals can be compared visually with those for P. The S-wave residuals relative to P increase sharply northwards of the ITS by 1.5 s (except at Lhasa, station K). Such a strong relative spatial variation cannot be accounted for by variation in layer thickness of different solid rocks as their seismic-wave velocity ratios do not differ much, and are close to 1.8. This discrepancy of P and S behaviour over a short distance in the array supports the inference that it is caused by partial melt in the crust or uppermost mantle. Elevation-corrected waveforms are shown in *c* for PKP waves and in *d* for SKS waves plotted in time as residuals relative to standard tables. Ratio of timescales is 1.8. The differential variation (*a*, *b*) for one wave with respect to the other along the array can be appreciated directly in the record section images, without onset picking.

Maxiang–Yangbajain, the relative delay increases strongly (Figs 1 and 3, stations N, M, L); in contrast, the splitting time remains small at first and then increases to one second in the north of the Lhasa block (Fig. 1). These results could be explained by crustal melts in the Yangbajain geothermal region, which would be more effective in delaying S waves than causing splitting.

Further northeast, the larger splitting with comparable S delay is more readily explained by flow in the mantle. The fast directions that we observe will in either case give the horizontal orientation of shear or flow, which for fluid-filled defects could either be in the crust and mantle, but for oriented olivines would be confined to the mantle. The fast directions need not be aligned with the trend of upper-crustal deformation or the general motion of crustal plates, as the ductile flow can be decoupled from the brittle deformation. This is what we observe, because fast directions do not align directly with the local orientation of shear zones at the surface, such as the KJFZ; however, the ENE orientation of mantle flow is consistent with an averaged model of upper-crustal flow. Taking a broad view over central Asia, Avouac and Tapponnier¹³ measured slip at the main fault zones and proposed a four-plate kinematic model of rigid-block rotations of the crust. They derived a regional velocity field with respect to stable Siberia in which the motion in the region where the KJFZ and BNS cross our profile approaches the 70° E orientation of seismic anisotropy that we observe.

We find that the Tibet-Himalayas collision zone is divided into at least three zones with very different anisotropy parameters. To the south, on the India plate where collisional effects are expected to be small, the measured anisotropy orientations are northeast, parallel to ancient plate motion. Within the Tethyan Himalayas, the orientations are aligned with the north-northwest differential motion between India and Tibet; here anisotropy indicates imbrication of their lithospheres as the cause of crustal thickening and elevation. North of the ITS anisotropy orientation changes, and relative S-wave delays and splitting times are largest, presumably related to elevated temperatures in the crust and mantle, which is inconsistent with the further extension here of the cold cratonic lithosphere of India. The fast orientation indicates an east-northeastwards flow of the mantle which is discordant with surface deformation, as indicated by geological features separating crustal blocks. This discordance suggests that the brittle upper crust is broken into blocks with relative motion occurring along shear zones which are not parallel to mantle flow. When the block motions are summed, however, the direction of the net motion of the brittle upper crust agrees with the east-northeast orientation of deeper ductile flow indicated by seismic anisotropy. □

Record of emergent continental crust ~3.5 billion years ago in the Pilbara craton of Australia

Roger Buick*, J. R. Thorne[†], N. J. McNaughton*, J. B. Smith*, M. E. Barley* & M. Savage*

* Key Centre for Strategic Mineral Deposits, Department of Geology and Geophysics, University of Western Australia, Nedlands, Western Australia 6907, Australia

[†] Sipa Resources Ltd, PO Box 1183, West Perth, Western Australia 6872, Australia

ISOTOPIC data for the Earth's oldest rocks¹⁻⁷ imply that a considerable volume of continental crust existed during the early Archaean aeon (>3.0 Gyr ago), but it is not known when this crust first began to form emergent landmasses. Sedimentary geochemistry suggests^{8,9} that the area of exposed continent was negligible until late in the Archaean¹⁰, a contention supported by the fact that, until now, all greenstone supracrustal volcanic and sedimentary successions shown to have been deposited on eroded continental basement have yielded ages of ≤ 3.0 Gyr. Here we report the discovery of an angular unconformity (an ancient erosion surface) beneath rocks of the 3.46-Gyr Warrawoona Group in the Pilbara craton of Australia, currently the oldest known well-preserved greenstone succession. Below the unconformity, low-grade greenstones older than 3.5 Gyr were intruded by voluminous granitoids before erosion. As the overlying Warrawoona rocks are only mildly metamorphosed, slightly deformed and were deposited near sea level, we infer that they accumulated on crust that was already rigid, cool and buoyant. Thus, by 3.46 Gyr ago, the sub-Warrawoona rocks formed an emergent block of continental crust, the most ancient known.

The unconformity is best exposed in the Pilgangoora syncline between 21° 05' S, 119° 13' E and 21° 07' S, 118° 59' E, although it crops out over a strike length of at least 75 km (Fig. 1). Previously recognized as a minor disconformity^{12,13}, it is marked by angular discordances in bedding of up to 60° (Fig. 2a), but in most places the rocks above and below are subconcordant. The unconformity surface has at least 10 m of eroded topography, with underlying chert beds forming steep ridges that are draped by the overlying succession (Fig. 2c). The basal bed of the overlying succession consists of well sorted, well rounded, mature quartz-chert sandstone^{13,14} which grades into cherty conglomerate and breccia near the upstanding ridges (Fig. 2c). Similar sandstone fills fissures that penetrate up to 100 m into the underlying rocks. Beneath the unconformity, igneous rocks are markedly more altered within 50 m of the upper contact, felsics becoming more kaolinitic and mafics less chloritic, possibly representing a palaeosol. Depositional environment changes across the unconformity, the underlying rocks having formed under deep water, whereas those above were deposited at shallow depths. The unconformity is also marked by a change in metamorphic grade, from upper-greenschist facies below to lower-greenschist above. The lateral change in concordance between the upper and lower successions indicates that the lower rocks were folded before formation of the unconformity. In places, the unconformable contact is perfectly exposed with no tectonic fabric evident, precluding juxtaposition along a fault or shear.

Rocks above the unconformity can be correlated lithostratigraphically with the upper part of the Warrawoona Group in the nearby North Pole Dome¹⁵ (Fig. 3). The succession consists of alternating units up to 1 km thick of ocellar magnesian basalt (7-11% MgO) and subophitic tholeiitic basalt (3.5-7% MgO), each with abundant pillowed and coarsely amygdaloidal (0.5-

Received 14 July 1993; accepted 15 May 1995.

- Ni, J. & Barazangi, M. *J. geophys. Res.* **89**, 1147-1163 (1984).
- Molnar, P. *Phil. Trans. R. Soc. A* **326**, 33-88 (1988).
- England, P. & Houseman, G. *J. geophys. Res.* **94**, 17561-17579 (1989).
- Hirn, A. et al. *Nature* **307**, 25-27 (1984).
- Le Pichon, X., Fournier, M. & Jolivet, L. *Tectonics* **11**, 1085-1098 (1992).
- Armijo, R., Tapponnier, P. & Han, T. *J. geophys. Res.* **94**, 2787-2838 (1989).
- Silver, P. & Chan, W. W. *J. geophys. Res.* **96**, 16429-16454 (1991).
- Vinnik, L. P., Makeyeva, L. I., Milev, A. & Usenko, A. Y. *Geophys. J. Int.* **111**, 433-447 (1992).
- Nicolas, A. & Christensen, N. I. in *Composition, Structure and Dynamics of the Lithosphere—Asthenosphere System* (eds Fuchs, K. & Froidevaux, C.) 111-123 (Geodyn. Ser. 16, Am. geophys. Un., Washington DC, 1987).
- Makeyeva, L. I., Vinnik, L. P. & Roecker, S. W. *Nature* **358**, 144-147 (1992).
- Gao, S. et al. *Nature* **371**, 149-151 (1984).
- Molnar, P. *Nature* **358**, 105-106 (1992).
- Avouac, J. C. & Tapponnier, P. *Geophys. Res. Lett.* **20**, 895-898 (1993).
- Gripp, A. E. & Gordon, R. G. *Geophys. Res. Lett.* **17**, 1109-1112 (1990).
- McNamara, D. E., Owens, T. J., Silver, P. G. & Wu, F. T. *J. geophys. Res.* **99**, 13655-13665 (1994).
- Hirn, A. et al. *Nature* **307**, 23-25 (1994).
- Barazangi, M. & Ni, J. *Geology* **10**, 179-185 (1982).
- Brandon, C. & Romanowicz, B. *J. geophys. Res.* **91**, 6547-6564 (1986).
- Bourjot, L. & Romanowicz, B. *Geophys. Res. Lett.* **19**, 881-884 (1992).
- Molnar, P. *Earth planet. Sci. Lett.* **101**, 68-77 (1990).
- Mavko, G. M. *J. geophys. Res.* **85**, 5173-5189 (1990).
- Wu, G. J., Xiao, X. C. & Li, T. D. *Global Geoscience Transect Vol. 3* (Am. Geophys. Un., Washington DC, 1991).
- Coulon, C., Maluski, H., Bollinger, L. & Wang, S. *Earth planet. Sci. Lett.* **79**, 281-302 (1986).
- Turner, S., Sandiford, M. & Foden, J. *Geology* **20**, 931-934 (1992).
- Molnar, P. & Chen, W. P. *J. geophys. Res.* **88**, 1180-1196 (1983).

ACKNOWLEDGEMENTS. We thank P. Davis, R. Girdler & J. McBride for helping to improve the manuscript. The French-Chinese experiment was supported by the INSU of Centre National de la Recherche Scientifique and Direction de la Recherche et des Etudes Doctorales, Ministry of Public Education, Paris and by the Chinese Academy of Geological Sciences, Ministry of Geology, Beijing.

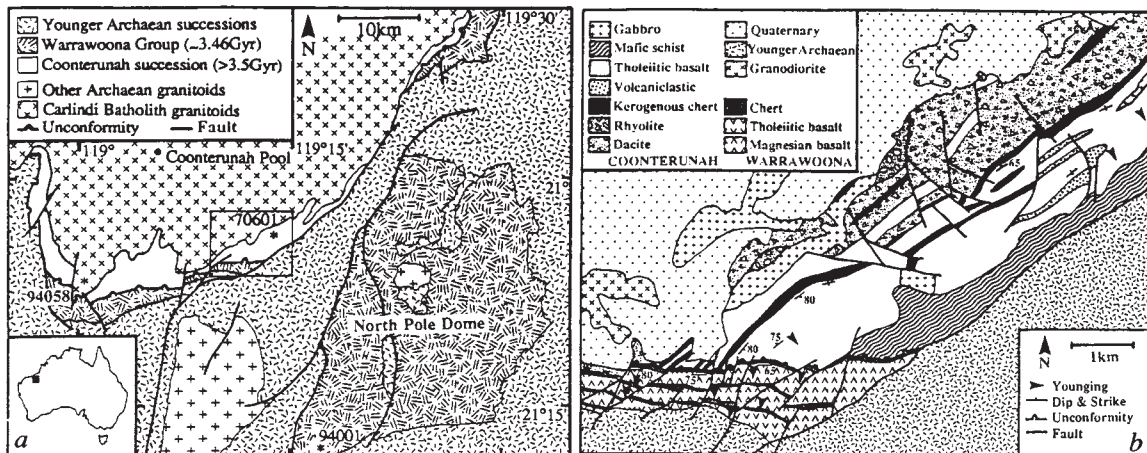


FIG. 1 a, Map showing the location and geological relationships of the Coonterunah succession and its overlying unconformity, box shows

location of Fig. 1b; b, Detailed map of part of the Coonterunah–Warrawoona unconformity.

2 cm) flows, separated by thin (<50 m) chert beds. The chert immediately above the unconformity, the “Strelley Pool Chert”¹⁴, has a distinctive internal stratigraphy with interlayered beds of partly silicified wavy-laminated or acicular-crystalline dolostone capped by silicified basaltic sand or silty mafic ash with pseudomorphs after diagenetic gypsum crystals. Higher cherts mostly consist of the latter lithology with some correlatable interbeds of accretionary lapilli arenite¹⁶. The depositional setting was shallow-water, indicated by the abundance of carbonate and sulphate evaporites, phreatomagmatic tephra and coarsely amygdaloidal basalts^{14,16}. Mineral assemblages (chlorite–albite–epidote in tholeiites) indicate very-low-grade metamorphism of prehnite–pumpellyite to lower-greenschist facies under very-low-strain conditions. Deformation has been restricted to steep tilting and small-scale block faulting, with late strike-slip faults truncating large tracts of relatively undisturbed stratigraphy.

Beneath the unconformity, the rocks belong to a new stratigraphic unit, here called the Coonterunah succession after a local landmark (Fig. 1). This succession, at least 5.5 km thick, is dominated by a bimodal suite of volcanic rocks with compositional units ~1 km thick (Fig. 3). Mafic units consist of pillowed and massive, sparsely amygdaloidal, tholeiitic basalt with sub-

ordinate magnesian basalt, komatiitic basalt (>11% MgO) and coarse-grained volcaniclastic rocks, all intruded by voluminous gabbro stocks and sills. Felsic units contain massive, highly amygdaloidal dacite with subordinate dacitic porphyry and hyaloclastic rhyolite, all intruded by gabbro. Volcanic units are commonly separated by thin (1–20 m) beds of plane-laminated kerogenous chert, ironstone and dolostone. The depositional environment was deep-water, shown by the presence of pillows in sparsely amygdaloidal basalts, the absence of breccia in highly amygdaloidal dacites and the scarcity of sedimentary bedforms and detrital clastic sediment. The succession is intruded at its base by regionally extensive granitoids of the Carlindi batholith, producing a hornfels aureole several hundred metres thick. Elsewhere, Coonterunah rocks have mostly been metamorphosed to middle- to upper-greenschist facies (biotite–muscovite assemblages in pelites) with slight schistosity, with higher grades only attained under high-strain conditions in the far west of the outcrop area. Deformation is similar to that displayed by the rocks above the unconformity, with the addition of an early episode of open folding.

In one area (21° 08' S, 119° 00' E), the unconformity cuts both volcanic rocks of the Coonterunah succession and granitoids of the Carlindi batholith. As elsewhere, the granitoids intrude

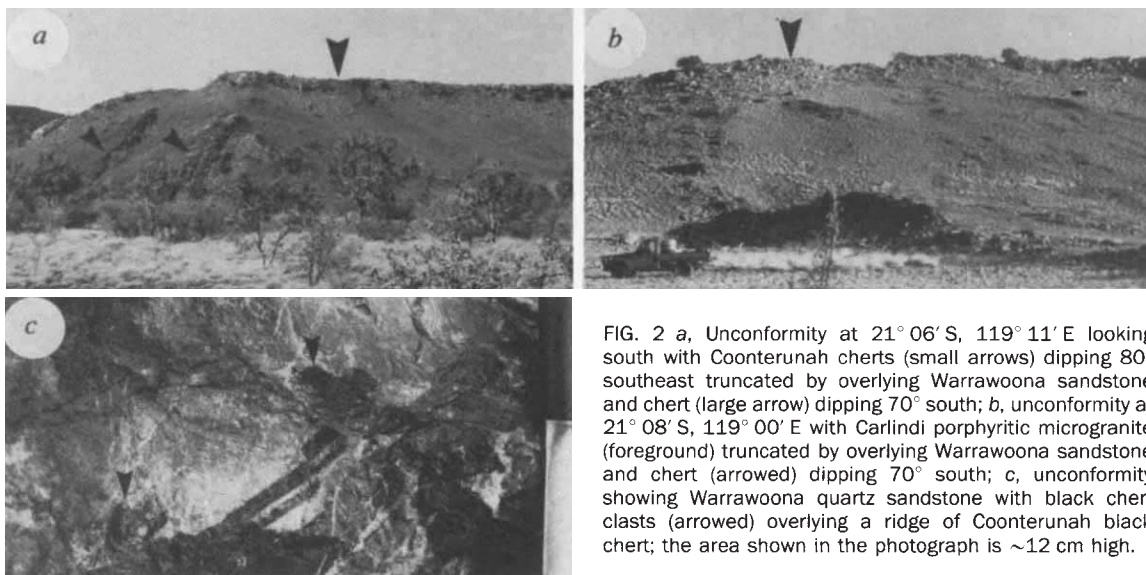


FIG. 2 a, Unconformity at 21° 06' S, 119° 11' E looking south with Coonterunah cherts (small arrows) dipping 80° southeast truncated by overlying Warrawoona sandstone and chert (large arrow) dipping 70° south; b, unconformity at 21° 08' S, 119° 00' E with Carlindi porphyritic microgranite (foreground) truncated by overlying Warrawoona sandstone and chert (arrowed) dipping 70° south; c, unconformity showing Warrawoona quartz sandstone with black chert clasts (arrowed) overlying a ridge of Coonterunah black chert; the area shown in the photograph is ~12 cm high.

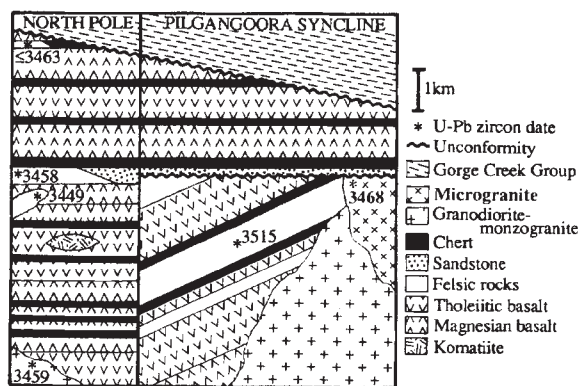


FIG. 3 Schematic stratigraphic diagram showing (left) the Warrawoona Group in the North Pole Dome; and (right) the Coonterunah succession overlain by the Warrawoona Group and intruded by the Carlindi batholith in the Pilgangoora syncline. U-Pb zircon ages (Myr) are from this study and ref. 19.

Coonterunah rocks, but here the nonfoliated granodioritic pluton is itself intruded by massive porphyritic microgranite containing β -quartz and orthoclase phenocrysts. This late intrusive episode occurred before the unconformity formed, as the microgranite is truncated by the erosion surface (Fig. 2b) with no change in grain size or texture approaching the contact. That such a fractionated, high-level microgranite should intrude an extant granitoid-greenstone terrain suggests that emplacement occurred after peak metamorphism and deformation, when little erosion was required to generate the unconformity.

Zircon U-Pb geochronology constrains the age of rocks below and above the unconformity (Fig. 3). A brecciated hyaloclastic rhyolite from the middle of the Coonterunah succession (sample no. 70601) yields abundant clear to pale-brown, euhedral to subhedral zircons with faint euhedral zoning and no distinct core. A number of grains (27) were analysed by SHRIMP II mass spectrometry using established methods^{17,18} and isotopic constants¹⁹. The data, all concordant to near-concordant, have the same $^{207}\text{Pb}/^{206}\text{Pb}$ within experimental error (Fig. 4a). Hence,

we regard the weighted $^{207}\text{Pb}/^{206}\text{Pb}$ age for all data, $3,515.2 \pm 2.7$ Myr, as the eruptive age of the rock. The Carlindi porphyritic microgranite (no. 94058), the youngest intrusive rock predating the unconformity, contains a morphologically homogeneous population of euhedral, pale-brown, finely-zoned zircons, a near-concordant subset of which (Fig. 4b) yields a $^{207}\text{Pb}/^{206}\text{Pb}$ age of $3,467.6 \pm 3.7$ Myr. This probably represents the intrusive age of the rock, though if there were early Pb loss, the ~ 3.50 -Gyr age of two older grains without obvious xenocrystic features may provide a better minimum estimate. In the succession immediately above the unconformity, no suitable rocks for dating exist, but the correlative part of the Warrawoona Group in the nearby North Pole Dome contains tuffaceous sandstone of felsic origin ~ 3 km above the stratigraphic level of the unconformity. This rock (no. 94001), a rippled pumiceous arenite with minor clastic quartz, muscovite and chert, has a diverse suite of zircons, many of which are rounded and detrital. Most grains yield ages similar to the Coonterunah sample, but some have more typical Warrawoona ages^{20,21} of 3.48–3.46 Gyr (Fig. 4c). The best interpretation of these data is that the older zircons are inherited from Coonterunah and Carlindi rocks, whereas the youngest concordant zircon, an angular fragment, records an eruption age of $3,463 \pm 7$ Myr for the juvenile felsic material. However, it is also possible that all of the zircons are detrital and that the depositional age of the sediment is somewhat younger.

Thus, the unconformity most probably formed between 3.47 and 3.46 Gyr, by which time the Carlindi-Coonterunah granitoid-greenstone terrain beneath it had evolved into a relatively rigid, cool and buoyant crustal block. This functioned as continental crust during deposition of the overlying Warrawoona rocks, as they have never been strongly deformed, have only been subjected to the mildest metamorphism and were deposited in considerable thickness over a wide area at shallow water depths. It already had a continental composition when the Warrawoona rocks formed, as the dominantly basaltic-dacitic Coonterunah greenstones were intruded by voluminous granitoids before the unconformity developed, producing a sialic terrain that shed quartzose sediment to form the bed of mature

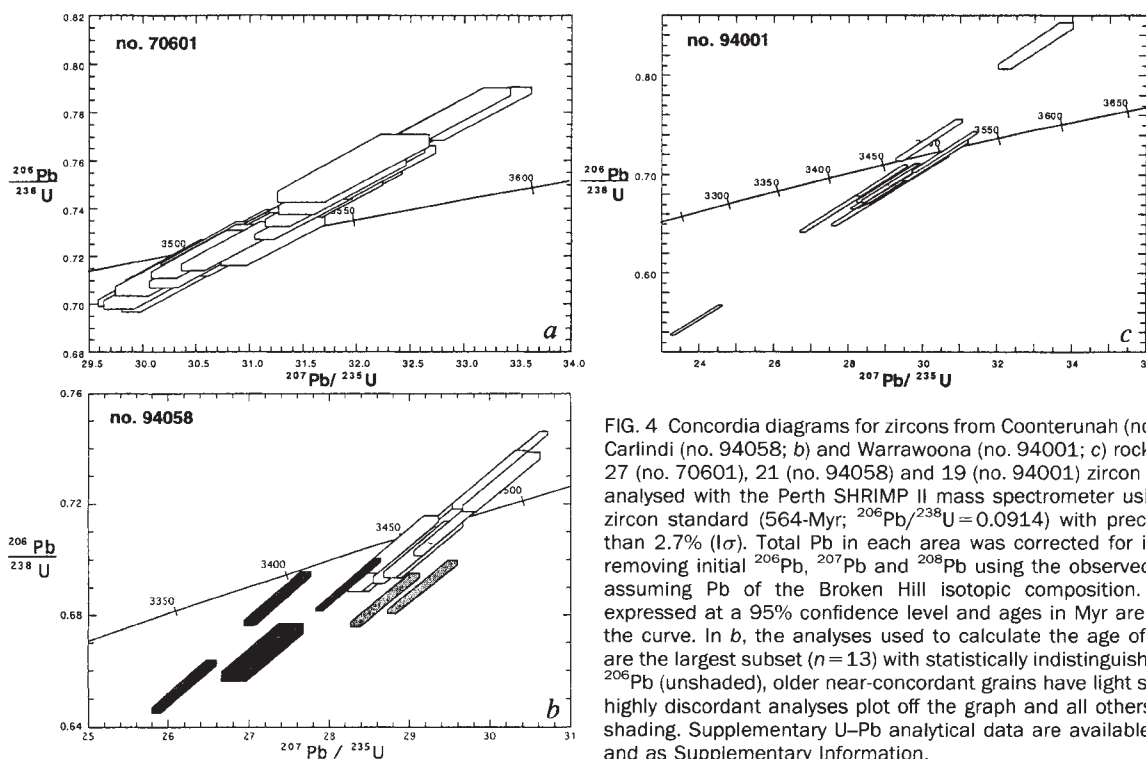


FIG. 4 Concordia diagrams for zircons from Coonterunah (no. 70601; a) Carlindi (no. 94058; b) and Warrawoona (no. 94001; c) rocks. Suites of 27 (no. 70601), 21 (no. 94058) and 19 (no. 94001) zircon grains were analysed with the Perth SHRIMP II mass spectrometer using the CZ3 zircon standard (564-Myr; $^{206}\text{Pb}/^{238}\text{U} = 0.0914$) with precision better than 2.7% (1σ). Total Pb in each area was corrected for initial Pb by removing initial ^{206}Pb , ^{207}Pb and ^{208}Pb using the observed ^{204}Pb and assuming Pb of the Broken Hill isotopic composition. Errors are expressed at a 95% confidence level and ages in Myr are marked on the curve. In b, the analyses used to calculate the age of no. 94058 are the largest subset ($n = 13$) with statistically indistinguishable $^{207}\text{Pb}/^{206}\text{Pb}$ (unshaded), older near-concordant grains have light shading, two highly discordant analyses plot off the graph and all others have dark shading. Supplementary U-Pb analytical data are available from R. B. and as Supplementary Information.

sandstone immediately above the unconformity. Despite evidence suggestive of a continental basement to the Warrawoona Group (~3.7-Gyr zircon xenocrysts in Warrawoona volcanics²⁰, ~3.3-Gyr Pilbara granitoids derived from much older crust²²), the discovery of a preserved substrate is unexpected, as recent reviews regard the Warrawoona Group as oceanic or arc-related^{23,24}. Indeed, stratigraphic proof of continental basement beneath greenstone belts older than ~3.0 Gyr has hitherto proved elusive¹¹. Hence, the discovery of a sub-Warrawoona unconformity extends the record of emergent continental blocks back another 0.5 Gyr, well into the early Archaean. This finding thus supports tectonic evolutionary models implying that substantial volumes of continental crust existed during the first billion years of the Earth's history^{10,25}. □

Received 1 February; accepted 15 May 1995.

- Jacobsen, S. B. & Dymed, R. F. *J. geophys. Res.* **93**, 338–354 (1988).
- Bowring, S. A., King, J. E., Housh, T. B., Isachsen, C. E. & Podosek, F. A. *Nature* **340**, 222–225 (1989).
- Bennett, V. C., Nutman, A. P. & McCulloch, M. T. *Earth planet. Sci. Lett.* **119**, 299–317 (1993).
- Froude, D. O. *et al.* *Nature* **304**, 616–618 (1983).
- Compston, W. & Pidgeon, R. T. *Nature* **321**, 766–769 (1986).
- Myers, J. S. *Precamb. Res.* **38**, 297–307 (1988).
- Nutman, A. P. & Collerson, K. D. *Geology* **19**, 791–794 (1991).

- Taylor, S. R. & McLennan, S. M. *The Continental Crust: its Composition and Evolution* (Blackwell, Oxford, 1985).
- Stevenson, R. K. & Patchett, P. J. *Geochim. cosmochim. Acta* **54**, 1683–1697 (1990).
- McCulloch, M. T. & Bennett, V. C. *Geochim. cosmochim. Acta* **58**, 4717–4738 (1994).
- Bickle, M. J., Nisbet, E. G. & Martin, A. J. *Geol.* **102**, 121–138 (1994).
- DiMarco, M. J. & Lowe, D. R. *Precamb. Res.* **44**, 147–169 (1989).
- DiMarco, M. J. & Lowe, D. R. *Sedimentology* **36**, 821–836 (1989).
- Lowe, D. R. *Precamb. Res.* **19**, 239–283 (1983).
- Hickman, A. H. *Bull. geol. Surv. W. Aust.* **127**, 1–167 (1983).
- Buick, R. & Barnes, K. R. *Publ. Geol. Dept. Extension Service Univ. West. Aust.* **9**, 37–53 (1984).
- Compston, W., Williams, I. S. & Meyer, C. J. *geophys. Res. Suppl.* **89**, B525–B534 (1984).
- Williams, I. S. & Claessens, S. *Contr. Miner. Petrol.* **97**, 205–217 (1987).
- Jaffrey, A. H., Flynn, K. F., Giendennin, L. E., Bentley, W. C. & Essling, A. M. *Phys. Rev. C* **4**, 1889–1906 (1971).
- Thorpe, R. I., Hickman, A. H., Davis, D. W., Mortensen, J. K. & Trendall, A. F. *Precamb. Res.* **56**, 169–189 (1992).
- McNaughton, N. J., Compston, W. & Barley, M. E. *Precamb. Res.* **60**, 69–98 (1993).
- Bickle, M. J. *et al.* *Precamb. Res.* **60**, 117–149 (1993).
- Krapez, B. *Precamb. Res.* **60**, 1–45 (1993).
- Barley, M. E. *Precamb. Res.* **60**, 47–67 (1993).
- Jacobsen, S. B. *Earth planet. Sci. Lett.* **90**, 315–329 (1988).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. We thank the Sipa-Outokumpu Panorama Project for access to unpublished data, M. Dahl, L. Wang and I. R. Fletcher for geochronological assistance, D. I. Groves and A. H. Knoll for criticism, and M. J. Bickle and P. Hoffman for reviews. Zircon analyses were performed on a sensitive high-resolution ion microprobe mass spectrometer (SHRIMP II) operated by a consortium consisting of Curtin University of Technology, the Geological Survey of Western Australia and the University of Western Australia with the support of the Australian Research Council. This work was funded by an Australian Research Fellowship (R.B.) and by an ARC grant (M.E.B.).

Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation

Debabrata Mukhopadhyay*, Leonidas Tsiokas*, Xiaoli Zhou†, David Foster‡, Joan S. Brugge† & Vikas P. Sukhatme*§

* Beth Israel Hospital and Harvard Medical School, DANA 517, 330 Brookline Avenue, Boston, Massachusetts 02215, USA.

† ARIAD Pharmaceuticals, Cambridge, Massachusetts 02139, USA

‡ Hunter College, New York, New York 10018, USA

ANGIOGENESIS, the formation of new microvasculature by capillary sprouting, is crucial for tumour development¹. Hypoxic regions of solid tumours produce the powerful and directly acting angiogenic protein VEGF/VPF (vascular endothelial growth factor/vascular permeability factor)^{2–6}. We now investigate the signal transduction pathway involved in hypoxic induction of VEGF expression. Hypoxia is known to induce a tyrosine kinase cascade that results in the activation of nitrogen-fixation genes in *Rhizobium meliloti*⁷, and activation of tyrosine kinases is critical in signalling triggered by growth factors and ultraviolet light. We show here that genistein, an inhibitor of protein tyrosine kinase⁸, blocks VEGF induction. Hypoxia increases the kinase activity of pp60^{c-src} and its phosphorylation on tyrosine 416 but does not activate Fyn or Yes. Expression of either a dominant-negative mutant form of c-Src or of Raf-1 markedly reduces VEGF induction. VEGF induction by hypoxia in c-Src(–) cells is impaired, although there is a compensatory activation of Fyn. Our results provide an insight into hypoxia-triggered intracellular signalling, define VEGF as a new downstream target for c-Src, and suggest a role for c-Src in promoting angiogenesis.

A time course of VEGF mRNA induction in U87 glioma cells subjected to hypoxic conditions showed maximal induction at 4 hours (data not shown). Genistein led to a dose-dependent

decrease (half-maximal inhibitory concentration, IC₅₀ ≈ 36 μM) in hypoxia-induced VEGF messenger RNA expression (Fig. 1a). At 360 μM genistein, VEGF mRNA was reduced to basal levels. These concentrations are within the range that inhibits specific tyrosine kinases in other cell types^{9,10}. Cells treated with the highest concentration of genistein appeared morphologically normal, their RNA was not degraded, and the levels of other mRNAs, such as that for the transforming growth factor TGF-β1, were unaffected (Fig. 1a). Similar results were obtained in kidney 293 cells (Fig. 1b), in HT1080 human fibrosarcoma cells and in a renal carcinoma cell line (data not shown).

As genistein is an inhibitor of pp60^{c-src} kinase activity⁸, we investigated whether c-Src could be activated by hypoxia. Under hypoxic conditions, a time-dependent stimulation of Src kinase activity (using peptide A as substrate¹¹) was observed in U87 cells, with a maximal fourfold increase at 30 min (Fig. 2a). Activ-

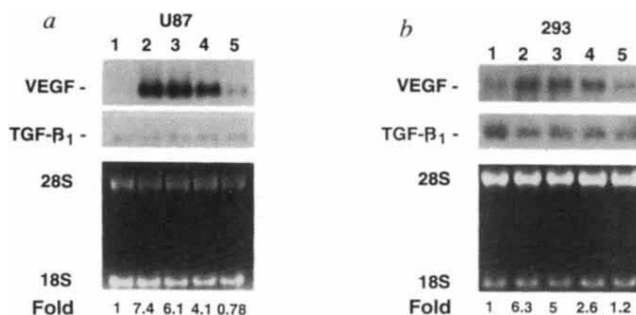


FIG. 1 Inhibition of hypoxia-induced VEGF mRNA expression by genistein. Total RNA (5 μg) was extracted from confluent U87 cells (ATCC HTB-14; a), and 293 cells (ATCC CRL 1573; b) incubated in normoxic (lane 1) or hypoxic (lanes 2–5) conditions (95% N₂ and 5% CO₂) in the absence (lane 2) or presence of 3.6 μM (lane 3), 36 μM (lane 4) or 360 μM genistein (lane 5). Blots were hybridized to human VEGF and TGF-β1 (for normalization) cDNA probes. Cells were preincubated with genistein for 30 min and then subjected to hypoxic conditions for 4 h. Lower panels, ethidium bromide-stained RNA samples before transfer; ribosomal RNA markers (18S and 28S) are indicated. Results were similar from two experiments.

§ To whom correspondence should be addressed.

ity on peptide B was ~15-fold less but showed a parallel time course, as expected, and peptide C was not phosphorylated; results were similar with 293 cells (Fig. 2b). The extent of c-Src activation in both cell types was comparable to that achieved with other stimuli of pp60^{c-src} (refs 12, 13). The relatively late induction of pp60^{c-src} activity (between 30 and 60 min) could be due to the fact that hypoxic conditions were not fully established in the culture medium until about 30 min after placement in the hypoxic environment (data not shown). This activation of c-Src by hypoxia is in agreement with previous results¹⁴. Figure 2c shows that in both U87 and 293 cells, hypoxia does not affect endogenous c-Src protein levels. As there is a correlation between phosphorylation of Tyr 416 and activation of Src^{15,16}, we assessed the phosphorylation of Src at Y416 using an antibody reactive against Src Y416 but not against non-phosphorylated Y416 Src or other phosphorylated tyrosines (Fig. 2c). Figure 2c shows that phosphorylation of Src Y416 is enhanced over a 60-min period of hypoxia, indicating that hypoxia increases the catalytic activity of Src as well as tyrosine phosphorylation of the Src autophosphorylation site.

To investigate whether there is a link between Src activation and VEGF mRNA induction, we introduced a c-src expression vector into cells and determined the levels of VEGF mRNA under hypoxic and normoxic conditions. Subconfluent cell cultures were used to maximize transfection efficiency; under these

conditions, hypoxia-induced VEGF expression is less than with confluent cultures. Figure 2d and e shows that c-Src overexpression causes a concentration-dependent increase (≤ 3 –4-fold) of VEGF transcripts under hypoxic conditions in both cell types. Figure 2f (lanes 3 and 4) shows that c-Src expression increases ~4-fold in cells transfected with 2 μ g of expression vector. Basal non-hypoxic levels of VEGF were notably unaffected (Fig. 2f, lanes 1 and 3), indicating that the increase in VEGF mRNA levels that was secondary to hypoxia is caused by activation of c-Src, not merely by overexpression of c-Src itself. We therefore investigated whether v-Src, a constitutively activated form of Src, could induce VEGF expression in the absence of hypoxia and found that transfection with a v-src expression vector increased VEGF mRNA to levels comparable with those induced by hypoxic conditions (Fig. 2f, lanes 5–8). Together these findings indicate that VEGF is a downstream target for activated Src.

To investigate whether c-Src activation is obligatory for VEGF induction by hypoxia, we analysed the hypoxic response in NIH3T3 cells stably transfected with an inducible c-src gene encoding a kinase-inactive form of Src. The ATP-binding site in this c-src mutant is inactivated by mutation of lysine 295 to arginine (K295R), and a phenylalanine substitution for tyrosine 527 (Y527F) prevents the intramolecular interaction between phosphorylated Y527 and the Src SH2 domain, rendering the

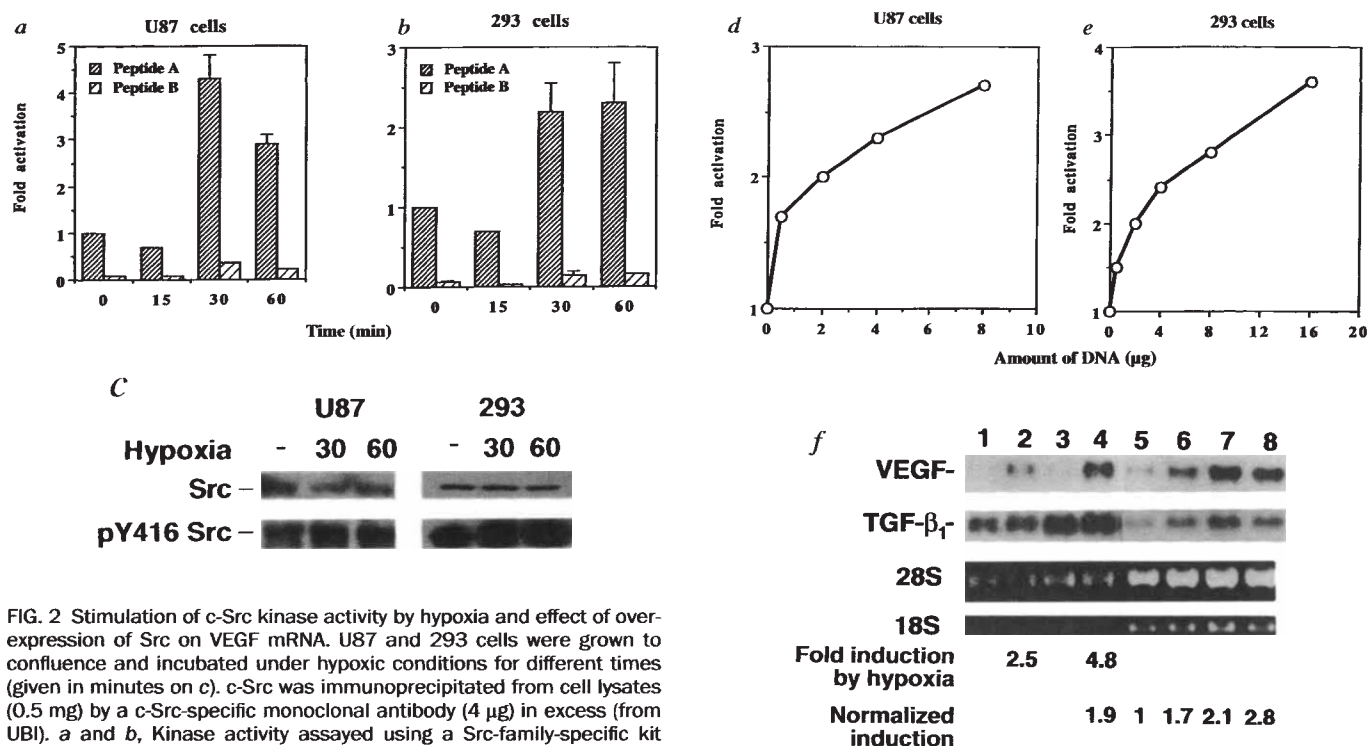


FIG. 2 Stimulation of c-Src kinase activity by hypoxia and effect of overexpression of Src on VEGF mRNA. U87 and 293 cells were grown to confluence and incubated under hypoxic conditions for different times (given in minutes on c). c-Src was immunoprecipitated from cell lysates (0.5 mg) by a c-Src-specific monoclonal antibody (4 μ g) in excess (from UBI). a and b, Kinase activity assayed using a Src-family-specific kit (UBI) containing three peptides: peptide A with Tyr 19 in Cdc2 (6–20-NH₂) replaced by Lys is an efficient substrate for Src kinases¹¹; peptide B, containing Val and Ser instead of Glu 12 and Thr 14, respectively, is 10–15-fold less efficiently phosphorylated by Src kinases¹¹; and peptide C, a potential target for the Ser/Thr kinases, is not a substrate for Src kinases¹¹ and contains no tyrosine. Counts for peptide C were not above background (data not shown). Results are the average of three independent experiments. Error bars for the 15-min time point are not discernible on the ordinate scale. c, Western blot analysis with antiphosphotyrosine-416-Src antibody and c-Src monoclonal antibody as determined on split samples. d, U87, and e, 293 cells were transiently transfected with different amounts of MoMLV LTR c-src expression plasmid²⁶ and subjected to hypoxic conditions for 4 h; VEGF mRNA expression was quantified, normalized to TGF- β_1 levels, and activation determined by comparison to values from empty vector controls. f, U87 cells were transfected with 2 μ g empty vector (lane 1, no hypoxia; lane 2, hypoxia); with 2 μ g c-src vector (lane 3, no hypoxia; lane 4, hypoxia);

with 8 μ g empty vector (lane 5); or with 2, 4 and 8 μ g of v-src expression vector lanes 6, 7 and 8 respectively). All experiments were done at least twice.

METHODS. For c, the peptide RLIEDNEPYTARQGAK (sequence from the Src catalytic domain) was coupled to BSA and injected into rabbits. The resulting polyvalent serum was adsorbed sequentially with a non-phosphorylated version of the peptide and with free phosphotyrosine, then purified by binding to the phosphorylated peptide using an Affigel column (S. Iyer et al., manuscript in preparation). d–f, Cells were plated at $2\text{--}3 \times 10^5$ cells per 60-mm dish one day before transfection with expression plasmids using calcium phosphate precipitation. Transfection efficiency was normalized by assaying β -galactosidase activity using a β -galactosidase gene under control of the cytomegalovirus immediate early promoter as internal control. In our hands, the variation of transfection efficiency was about 10% between plates.

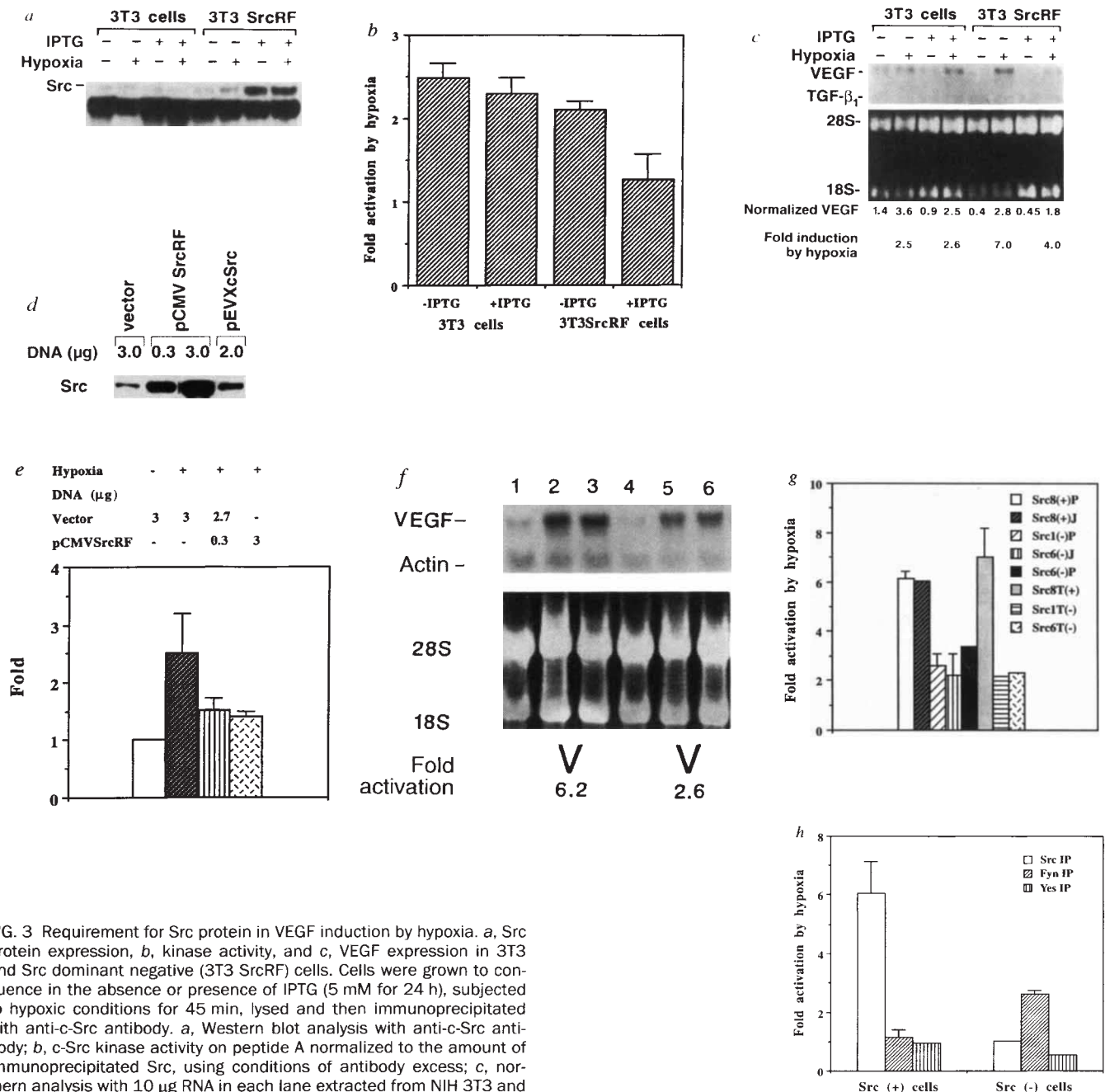


FIG. 3 Requirement for Src protein in VEGF induction by hypoxia. *a*, Src protein expression, *b*, kinase activity, and *c*, VEGF expression in 3T3 and Src dominant negative (3T3 SrcRF) cells. Cells were grown to confluence in the absence or presence of IPTG (5 mM for 24 h), subjected to hypoxic conditions for 45 min, lysed and then immunoprecipitated with anti-c-Src antibody. *a*, Western blot analysis with anti-c-Src antibody; *b*, c-Src kinase activity on peptide A normalized to the amount of immunoprecipitated Src, using conditions of antibody excess; *c*, Northern analysis with 10 µg RNA in each lane extracted from NIH 3T3 and 3T3 SrcRF cells incubated in normoxic or hypoxic conditions for 4 h in the absence or presence of IPTG; *d*, western blot analysis with c-Src antibody in 293 cells; *e*, 293 cells were transiently transfected with different amounts of CMV immediate early promoter-driven SrcRF expression plasmid and subjected to hypoxia for 4 h; VEGF mRNA was quantified and activation determined. Results are from three independent experiments. *f*, Inhibition of VEGF induction in *src*(-) cells. Northern analysis of 10 µg of RNA in each lane extracted from Src 8(+)/P cells (lanes 1–3) and Src 1(-)/P cells (lanes 4–6) incubated in normoxic (lanes 1, 4) or hypoxic conditions for 4 h (lanes 2, 3, 5, 6). Blots were hybridized with VEGF and β -actin cDNAs (for normalization). Lanes 2, 3 and 5, 6 represent independent dishes. *g*, Five independently derived Src(-) cell lines and three Src(+) lines were analysed. The 8T(+), 1T(-) and 6T(-) cells were immortalized with SV40 T antigen at passage 3 (ref. 18); the Src1(-)/P, Src6(-)/P (provided by P. Soriano), and Src6(-)/J (ref. 19) cell lines were derived by passage through crisis. Src8(+)/P were passage-69 cells; Src8(+)/J were early passage cells. Data with error bars are from at least three experiments; rest are the average of two independent experiments. *h*, Src8(+)/P cells and Src1(-)/P cells were incubated in hypoxic conditions for 45 min. Src, Fyn and Yes proteins were immunoprecipitated (IP) from cell lysates by their respective anti-

bodies and their kinase activity assayed. Results for Src and Fyn are from three experiments and are the average of two experiments for Yes. Paired *t*-test gives $P < 0.05$ for Fyn activity in hypoxic Src(+) and Src(-) cells.

METHODS. The mouse *c-src* gene containing the K295R and Y527F codon substitutions (SrcRF cDNA; M. Trahey, S. Thomas and J.S.B., unpublished results) was subcloned into vector pL14, which is a modified variant of the pL1-3 CAT plasmid²⁷. This variant contains 10 additional *lac* operators (X.-M. Zhou, unpublished results), and the hygromycin-B-resistance gene. To generate SrcRF 3T3 cells, NIH3T3 cells were transfected with 10 µg pL14SrcRF and 10 µg pHBAalp267, modified to encode the neomycin-resistance gene (A. Klippel and L. T. Williams, unpublished results). Subclones resistant to both G418 and hygromycin B were isolated and assayed for inducible SrcRF expression. The clone used here showed ~5-fold induction of SrcRF and very low leakage in the absence of 5 mM IPTG. This SrcRF cDNA was also subcloned into a CMV expression vector, pCB6+ (ref. 28), to give pCMVSrcRF.

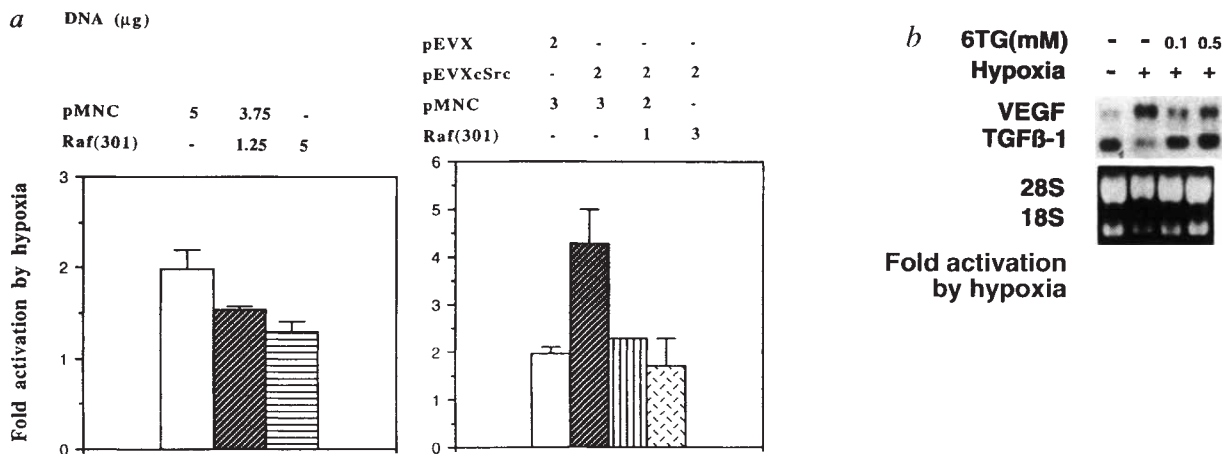


FIG. 4 Events downstream of c-Src activation. *a*, 293 cells were transfected with an expression vector containing the dominant interfering c-Raf-1 Raf(301) allele²⁰ with or without the MoMLV LTR c-Src expression plasmid pEVXcSrc. pMNC is the empty vector for Raf(301). Cells were subjected to hypoxia for 4 h; VEGF mRNA was quantified and fold

activation determined. Data for the 1 μg Raf(301) point are the average of two experiments; for all other points they are derived from three experiments. *b*, Total RNA (5 μg) was extracted from confluent U87 cells incubated under hypoxic conditions for 4 h in the presence or absence of 6-thioguanine (6TG) and fold activation calculated.

SH2 domain, and probably the SH3 domain as well, accessible to cellular binding proteins. This variant of Src (SrcRF) should act as a dominant negative protein by binding to cellular proteins that interact with activated variants of Src under normal conditions.

We measured the amount of Src protein (normal and dominant negative) in 3T3 cells and clone 3T3SrcRF with or without hypoxia in the presence or absence of isopropylthiogalactoside (IPTG). IPTG led to a ~5-fold increase (Fig. 3*a*) in 3T3SrcRF cells and no change in 3T3 cells. To investigate whether expression of this Src variant interfered with activity of endogenous Src, we tested the activity of Src during hypoxia in uninduced and induced (IPTG) conditions (Fig. 3*b*). The kinase activity of Src in 3T3 cells was elevated 2–3-fold by hypoxia in the absence of IPTG (Fig. 3*b*), extending our observation of hypoxia-induced Src activation in U87 and 293 cells. But after treatment of 3T3SrcRF cells with IPTG (Fig. 3*b*), there was a significant decrease in Src activity under hypoxic conditions (down to 1.3-fold). The increased induction of VEGF mRNA in 3T3 cells was unaffected by IPTG treatment (Fig. 3*c*). In uninduced (no IPTG) 3T3SrcRF cells, VEGF was increased 7.0-fold under hypoxic conditions (Fig. 3*c*) relative to the parental 3T3 cells), a result of the lower basal level of VEGF expression rather than of an increase in VEGF mRNA with hypoxia. After induction of SrcRF by IPTG, the hypoxic induction of VEGF was reduced 42% to 4.0-fold (Fig. 3*c*). The incomplete inhibition of the hypoxic response may reflect inadequate expression of SrcRF, or factors that compensate in the absence of Src-related kinases. We therefore transiently transfected a cytomegalovirus-driven SrcRF plasmid (pCMVSrcRF) into 293 cells and found high levels of SrcRF protein (Fig. 3*d*) and a dose-dependent decrease in VEGF activation with increasing dominant negative construct (Fig. 3*e*; maximum inhibition 73%). As the transfection efficiency of these cells is ~60–70% (data not shown), we conclude that the activation of the Src family of tyrosine kinases by hypoxia is probably critical in VEGF mRNA induction.

To determine the role of these kinases, we studied VEGF induction in fibroblasts derived from *c-src*-disrupted mice^{17–19} and found that VEGF mRNA induction due to hypoxia was 50–70% less in several lines of *c-src*(–) cells than in *c-src*(+) cells (Fig. 3*f, g*). This incomplete inhibition could be due to compensation by other Src family members in *c-src*(–) cells or by other cellular pathways. In support of the former, we find that Fyn (but not Yes) is activated preferentially by hypoxia

(2.6-fold) in *c-src*(–) cells but not in *c-src*(+) cells (Fig. 3*h*). Collectively, therefore, our data point to a critical role for c-Src.

We investigated events coupling c-Src activation to VEGF induction. As Raf-1 acts downstream of Src signalling¹³, we used a plasmid overexpressing the dominant inhibitory Raf-1 mutant Raf(301) (ref. 20). This construct inhibited hypoxia-induced endogenous VEGF expression in a dose-dependent manner (maximal inhibition ~70%; Fig. 4*a*). Moreover, Raf(301) blocked c-Src-activated VEGF expression under hypoxic conditions (Fig. 4*a*), indicating that Raf is acting downstream of c-Src. To identify events downstream of Raf-1, we used a purine analogue, 6-thioguanine, a potent inhibitor of extracellular signal-regulated kinases (ERKs)⁹ to demonstrate a dose-dependent inhibition of hypoxic VEGF induction (Fig. 4*b*), which suggests that ERKs or related kinases participate in this process.

Several other hypoxia-inducible genes, for example those encoding erythropoietin, PDGF β-chain, endothelin and angiotensin-converting enzyme, have been identified^{21,22}. Hypoxia also results in activation of NF-κB and further induction of *c-fos*, *c-jun*, *c-jun-B*, and *jun-D*^{23,24}, but no downstream target gene for these ubiquitous transcription factors has yet been identified. We have presented evidence that hypoxia activates c-Src and that VEGF is one of its downstream targets. Inhibitors of protein tyrosine kinases may have application in antiangiogenic therapy, perhaps through intracellular targeting²⁵. □

Received 19 January; accepted 24 April 1995.

1. Folkman, J. *Ann. Intern. Med.* **1**, 96–100 (1975).
2. Shweiki, D., Itin, A., Soffer, D. & Keshet, E. *Nature* **359**, 843–845 (1992).
3. Plate, K. H., Breier, G., Weich, H. A. & Risau, W. *Nature* **359**, 845–848 (1992).
4. Jaye, M. et al. *Science* **233**, 541–544 (1986).
5. Houck, K. A. et al. *Molec. Endocr.* **5**, 1806–1814 (1991).
6. Breier, G., Albrecht, U., Sterrer, S. & Risau, W. *Development* **114**, 521–532 (1992).
7. Gilles-Gonzalez, M. A., Ditta, G. S. & Helinski, D. R. *Nature* **350**, 170–172 (1991).
8. Akiyama, T. et al. *J. Biol. Chem.* **262**, 5592–5595 (1987).
9. Qui, M.-S. & Green, S. H. *Neuron* **9**, 705–717 (1992).
10. Nakajima, K. & Wall, R. *Molec. Cell. Biol.* **11**, 1409–1418 (1991).
11. Cheng, H.-C., Nishio, H., Hatase, O., Ralph, S. & Wang, J. H. *J. Biol. Chem.* **267**, 9248–9256 (1992).
12. Gould, K. L. & Hunter, T. *Molec. Cell. Biol.* **8**, 3345–3356 (1988).
13. Devary, Y., Gottlieb, R. A., Smeal, T. & Karin, M. *Cell* **71**, 1981–1991 (1992).
14. Koong, A. C. et al. *Cancer Res.* **54**, 5273–5279 (1994).
15. Jove, R. & Hanafusa, H. A. *Rev. Cell Biol.* **3**, 31–56 (1987).
16. Kmiecik, T. E., Johnson, P. J. & Shalloway, D. *Molec. Cell. Biol.* **8**, 4541–4546 (1988).
17. Soriano, P., Montgomery, C., Geske, R. & Bradley, A. *Cell* **64**, 693–702 (1991).
18. Thomas, J. E., Soriano, P. & Brugge, J. S. *Science* **254**, 568–571 (1991).
19. Thomas, J. E., Aguzzi, A., Soriano, P., Wagner, E. F. & Brugge, J. S. *Oncogene* **8**, 2521–2529 (1993).
20. Kolch, W., Heidecker, G., Lloyd, P. & Rapp, U. R. *Nature* **349**, 426–428 (1991).

21. Kourembanas, S., Hannan, R. L. & Faller, D. V. *J. clin. Invest.* **86**, 670–674 (1990).
22. Kourembanas, S., Marsden, P. A., McQuillan, L. P. & Faller, D. V. *J. clin. Invest.* **88**, 1054–1057 (1991).
23. Goldberg, M. A. & Schneider, T. J. *J. biol. Chem.* **269**, 4355–4359 (1994).
24. Webster, K. A., Discher, D. J. & Bishopric, N. H. *J. biol. Chem.* **268**, 16852–16858 (1993).
25. Uckun, F. M. *et al. Science* **267**, 886–891 (1995).
26. Johnson, P. L., Coussens, P. M., Danko, A. V. & Shalloway, D. *Molec. cell. Biol.* **5**, 1073–1083 (1985).
27. Labow, M. A., Baim, S. B., Shenk, T. & Levine, A. J. *Molec. cell. Biol.* **10**, 3343–3356 (1990).
28. Brewer, C. B. *Meth. Cell Biol.* **43**, 233–245 (1994).

ACKNOWLEDGEMENTS. We thank P. Soriano for Src(–) cells; J. Bolen for anti-Fyn and anti-Yes antibody; J. Abraham for the VEGF cDNA; U. Rapp for the *raf* dominant-negative plasmid; M. Trahey and S. Thomas for construction of mouse SrcRf cDNA; A. Klippel and L. Williams for neo-hygro Lap plasmids; A. Laudano for rabbit polyvalent serum against Src phosphoY416; L. C. Cantley, T. Roberts and N. Mukhopadhyay for suggestions; B. Teicher for help with the hypoxic chamber; and L. G. Cantley, L. C. Cantley, F. Epstein and I. Drummond for critically reviewing the manuscript. This work was supported in part by grants from the NIH to V.P.S. and J.S.B.

Persistent DDT metabolite *p,p'*-DDE is a potent androgen receptor antagonist

William R. Kelce*, Christy R. Stone*,
Susan C. Laws*, L. Earl Gray*,
Jon A. Kempainen† & Elizabeth M. Wilson†‡

* Reproductive Toxicology Branch, Developmental Toxicology Division, Health Effects Research Laboratory, US Environmental Protection Agency, Research Triangle Park, North Carolina 27711, USA
The Laboratories for † Reproductive Biology and the Departments of Pediatrics, and ‡ Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599, USA

THE increase in the number of reports of abnormalities in male sex development in wildlife and humans coincided with the introduction of 'oestrogenic' chemicals such as DDT (1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane) into the environment. Although these phenotypic alterations are thought to be mediated by the oestrogen receptor, they are also consistent with inhibition of androgen receptor-mediated events. Here we report that the major and persistent DDT metabolite, *p,p'*-DDE (1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene), has little ability to bind the oestrogen receptor, but inhibits androgen binding to the androgen receptor, androgen-induced transcriptional activity, and androgen action in developing, pubertal and adult male rats. The results suggest that abnormalities in male sex development induced by *p,p'*-DDE and related environmental chemicals may be mediated at the level of the androgen receptor.

The increasing incidence of certain human male reproductive tract abnormalities (such as cryptorchidism and hypospadias) have occurred during periods of increased exposure to and body burdens of 'oestrogenic' chemicals^{1,2}. These effects are postulated to be mediated by the oestrogen receptor (ER) because *o,p'*-DDT, an environmental oestrogen, binds the ER^{3,4} and induces oestrogenic effects (such as increased uterine weight) in female rats^{5,6}. The phenotypic effects induced in the male, however, are reminiscent of androgen blockade at the level of the androgen receptor (AR)^{7,8}, which is interesting given that oestrogens are well-known antagonists of AR androgen binding⁹. In this regard, the ability of five of these suspected environmental endocrine disruptors to bind AR was investigated using cell-free binding assays in rat ventral prostate cytosol as previously described¹⁰. The structure of several of these environmental chemicals is illustrated together with diethylstilboestrol and 17 β -oestradiol in Fig. 1A. In competitive AR binding assays using [³H]R1881 (a radiolabelled synthetic androgen), all five environmental chemicals showed dose-dependent competitive inhibition supporting a lack of strict ligand specificity that is characteristic of the AR⁹.

TABLE 1 Inhibitor concentrations necessary for 50% displacement or inhibition (IC₅₀) of AR and ER binding and 5 α -reductase enzymatic activity by several suspected environmental endocrine disruptors

Chemical name	AR binding IC ₅₀ (μ M)	ER binding IC ₅₀ (μ M)	5 α -reductase IC ₅₀ (μ M)
Chlordecone	125	3	200
<i>p,p'</i> -DDT	75	> 1000	> 1000
<i>p,p'</i> -DDE	5	> 1000	> 1000
<i>o,p'</i> -DDT	95	5	> 1000
<i>p,p'</i> -DDD	90	> 1000	> 1000
Hydroxylflutamide	0.5	> 1000	ND
Diethylstilboestrol	10	0.0008	50
17 β -Oestradiol	0.5	0.002	50

Note that IC₅₀ and K_i values are not equivalent; the former are dependent on ligand concentration. Competitive binding assays were done by incubating cytosolic ventral prostate extracts from 24-h castrate rats for 20 h at 4 °C using 3 nM [³H]R1881 in the presence and absence of increasing concentrations of radioinert inhibitors. Nonspecific binding of [³H]R1881 was assessed by adding 100-fold molar excess radioinert R1881. Competitive binding by ER was assessed by incubating uterine cytosolic extracts from immature (20-day-old) rats for 30 min at 30 °C followed by 30 min at 4 °C using 1 nM [³H]17 β -oestradiol in the presence and absence of increasing concentrations of radioinert inhibitors. Nonspecific binding of [³H]17 β -oestradiol was assessed by adding 100-fold molar excess of radioinert DES. Bound [³H] ligand was isolated using hydroxylapatite extraction and quantified in ethanol extracts by scintillation counting as previously described for AR and ER¹⁰. Inhibition of 5 α -reductase activity was assessed by determining the ability of increasing concentrations of inhibitor to reduce the conversion of [³H]testosterone to [³H]5 α -dihydrotestosterone and [³H]5 α -androstan-3 α ,17 β -diol by microsomes isolated from the adult rat caput/corpus epididymis. Substrates were separated from products and quantified by HPLC with on-line radiochromatography as previously described²⁹. None of the DDT-related chemicals examined were unstable or metabolized *in vitro* as determined by HPLC/UV diode array spectroscopy following methylene chloride extraction of the assay media (see legend to Fig. 2). The data shown for AR and ER are representative of at least three independent experiments, whereas those for 5 α -reductase are derived from two independent experiments. ND, Not determined.

The most potent competitor, *p,p'*-DDE, had an inhibition constant, K_i, of 3.5 μ M, similar to that determined for the synthetic oestrogen, diethylstilboestrol (4.6 μ M) and about 30 times weaker than the natural oestrogen, 17 β -oestradiol (Fig. 2). The binding of *p,p'*-DDE to the AR was surprising and of particular concern as *p,p'*-DDE levels bioaccumulate because of slow metabolic degradation and excretion^{11,12}. *o,p'*-DDT, *p,p'*-DDT, *p,p'*-DDD (not shown) and chlordecone (Kepone) were at least 12–20-fold less effective than *p,p'*-DDE (Fig. 2). The linear slope replots indicate competitive inhibition.

Competitive binding assays with [³H]R1881 and [³H]17 β -oestradiol were used to compare the relative binding affinity of these chemicals for AR and ER, respectively. Chlordecone and *o,p'*-DDT were 20–40-fold more effective as competitors of oestrogen binding to the ER than androgen binding to the AR (Table 1). *p,p'*-DDT, *p,p'*-DDD and *p,p'*-DDE bound the AR 14, 11 and 200 times more effectively than the ER, respectively (Table 1). The poor ER binding of these chemicals agrees with previous reports^{6,13–15} and raises the possibility that the AR, rather than the ER, is the site of hormonal blockade by persistent environmental pollutants such as *p,p'*-DDE.

The ability of these chemicals to block androgen action at the level of the AR was investigated by transient cotransfection of monkey kidney CV1 cells with the human AR expression vector and a mouse mammary tumour virus (MMTV) promoter-luciferase reporter vector that assays androgen-induced transcriptional responses to AR¹⁶. *p,p'*-DDE and the potent antiandrogen, hydroxylflutamide, were equally effective inhibitors of androgen-induced AR transcriptional activity. About half the transcriptional activity induced with 0.1 nM 5 α -dihydrotestosterone was

inhibited by 0.2 μM p,p' -DDE or hydroxyflutamide (Fig. 3A). o,p' -DDT, a minor component of technical grade DDT, inhibited 5α -dihydrotestosterone-induced AR transactivation by 50 per cent at concentrations of 1 μM , whereas p,p' -DDT, p,p' -DDD and chlordecone exhibited weaker inhibition at these higher concentrations (Fig. 3A).

Inhibition of androgen-induced AR transcriptional activity by p,p' -DDE was not the result of p,p' -DDE cytotoxicity, as there was no significant change in dexamethasone-induced human glucocorticoid receptor activation of the MMTV luciferase reporter vector at identical concentrations of p,p' -DDE (Fig. 3Ba). In addition, CV1 cells transfected with the truncated and constitutively active human AR1-660 expression vector¹⁶ did not exhibit alterations in transcription in the presence of p,p' -DDE (Fig. 3Bb). Additional studies at higher concentrations showed that o,p' -DDT and p,p' -DDT were cytotoxic at concentrations greater than 5 μM , whereas p,p' -DDD, diethylstilboestrol and chlordecone were cytotoxic at concentrations greater than 1 μM . Finally, p,p' -DDE failed to induce transcriptional activity in CV1 cells in the absence of androgen suggesting that p,p' -DDE lacks AR agonist activity (data not shown). The results indicate that p,p' -DDE is an AR antagonist with a potency nearly equivalent to that of the antiandrogen, hydroxyflutamide. Finally, p,p' -DDE appears not to alter male reproductive development or function by inhibiting the activity of 5α -reductase, as concentrations required for inhibition of 5α -reductase activity (Table 1)

were 200–5,000 times higher than those required for inhibition of AR androgen binding and transcriptional activity, respectively.

Fetal, pubertal and adult male Long-Evans hooded rats were exposed to p,p' -DDE in order to establish the ability of p,p' -DDE to inhibit AR action *in vivo*. In developmental studies, pregnant dams ($n=8$ per group) were gavaged with vehicle (corn oil) or 100 mg p,p' -DDE per kg from gestational days 14–18. The resulting male pups exhibited reduced anogenital distance at birth ($P<0.04$; litter-means analysis) and retained thoracic nipples when examined on postnatal day 13 (Fig. 1B), both of which are indicative of prenatal antiandrogen activity^{17,18}. To determine the effect of p,p' -DDE on androgen-induced pubertal maturation in male rats, weanling (21 day) male rats were treated with 100 mg p,p' -DDE per kg per day until after puberty (day 57). Treatment with p,p' -DDE significantly delayed the onset of puberty by 5 days compared to control rats ($P<0.005$; Fig. 1C). The onset of puberty was defined as the day the prepuce separated from the penis (day 43 in the control rats)¹⁹. This delay in puberty induced by p,p' -DDE is consistent with that reported for the known environmental antiandrogen, vinclozolin^{10,20} and is not confounded by retarded growth, as p,p' -DDE-exposed rats were significantly heavier ($P<0.005$) at preputial separation relative to control rats (Fig. 1C). p,p' -DDE treatment did not reduce serum testosterone levels in these pubertal rats, suggesting that the antiandrogenic effects of p,p' -DDE exposure were not confounded by the reported ability of DDT-related chemicals

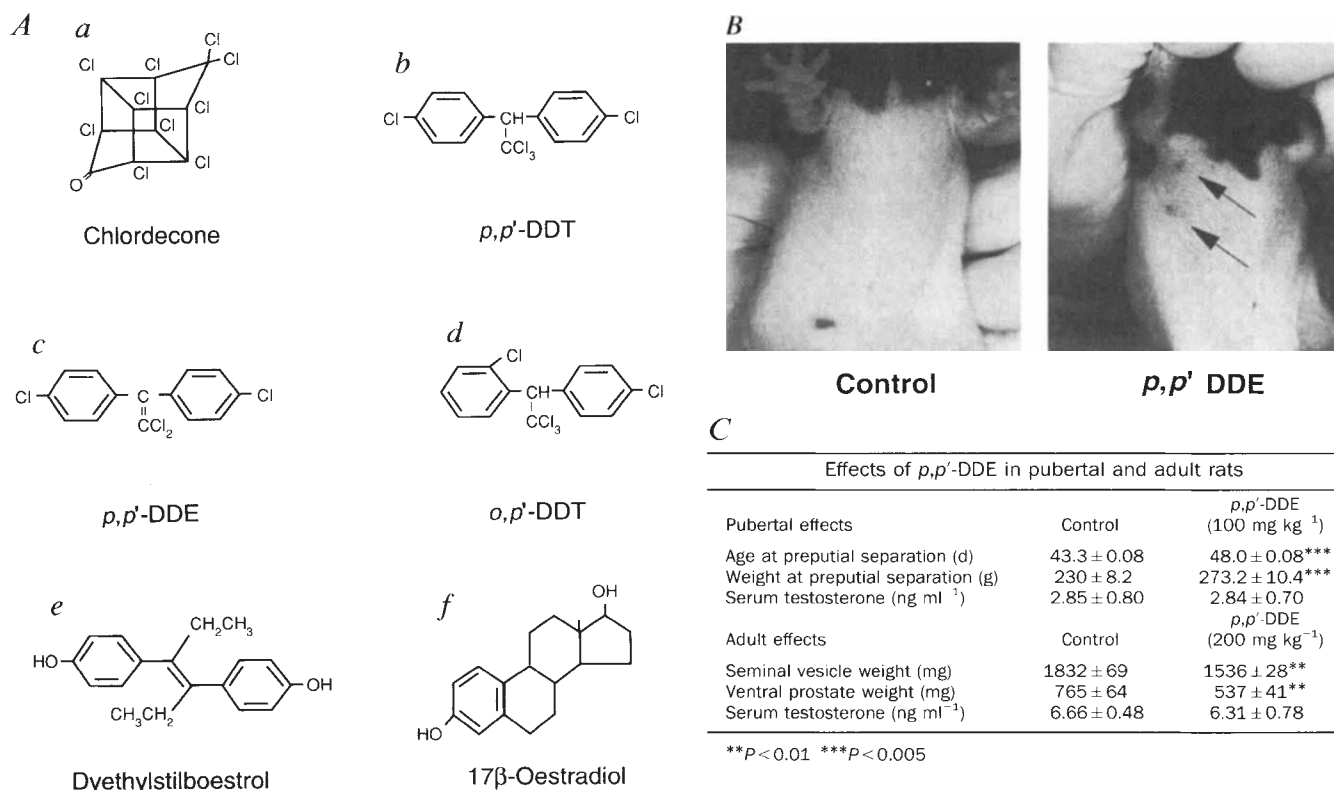


FIG. 1. A, Structural formulas of a, chlordecone (Kepone); b, p,p' -DDT; c, p,p' -DDE; d, o,p' -DDT; e, diethylstilboestrol; f, 17β -oestradiol. Note that technical DDT is a mixture of three forms, p,p' -DDT (85%), o,p' -DDT (15%) and o,o' -DDT (trace amounts). B, C, Antiandrogenic effects of p,p' -DDE in developing, pubertal and adult Long-Evans male rats. B, Pregnant rats were dosed daily by gavage with p,p' -DDE (100 mg per kg day) from gestational day 14–18. Males from the resulting litters typically displayed retention of thoracic nipples (arrows), an *in vivo* antiandrogenic anomaly not observed in control untreated rats. The rats were 13 days old at the time of the photographs. C, To establish the antiandrogenic effects of p,p' -DDE on male pubertal development, 25-day-old male rats ($n=12$ per group) were dosed daily by gavage with vehicle (corn oil) or p,p' -DDE (100 mg per kg per day) until day 57.

Treatment with p,p' -DDE delayed the onset of puberty (age at preputial separation) by 5 days compared to control rats ($P<0.005$) and was not associated with changes in body weight determined at the age of preputial separation or with altered circulating androgen levels. In the adult studies, 120-day-old male rats ($n=6$ per group) were castrated and implanted with testosterone-containing Silastic capsules (5 cm) to maintain constant serum androgen levels. Treatment with p,p' -DDE (200 mg per kg per day) by gavage daily for 4 days significantly reduced androgen-dependent seminal vesicle ($P<0.01$) and prostate ($P<0.01$) weights, compared to control vehicle-treated rats, despite high serum testosterone levels. Serum testosterone levels were not different between the two groups.

FIG. 2. Competitive inhibition of [3 H]R1881 binding to rat AR by a, chlordecone; b, *p,p'*-DDT; c, *p,p'*-DDE; d, *o,p'*-DDT; e, DES; and f, 17 β -oestradiol. A cell-free binding assay was done as described in Table 1. Data are presented as double-reciprocal plots together with the associated slope-replot analysis for the determination of apparent inhibition constants. The stability and metabolic fate of the DDT-related chemicals in each assay procedure was examined following methylene chloride extraction of the assay media by HPLC with in-line diode array detection. A Beckman System Gold (Solvent Module 126) HPLC equipped with a Rheodyne 7725i injector, a Rainin 100 μ l sample loop, a Rainin 4.6 $\times$ 250 mm Microsorb ODS reversed-phase C₁₈ column, and a Brownlee 3.2 $\times$ 15 mm MPLC reversed-phase C₁₈ guard column were used for the chromatography. Detection of *p,p'*-DDD, *p,p'*-DDT, *o,p'*-DDT and *p,p'*-DDE was accomplished by monitoring absorbance at 220 nm (with wavelength scans from 214–300 nm at 1-s intervals during the detection of each absorbance peak) with a Beckman 168 Diode Array Detector. An isocratic methanol:water (90:10) solvent system at a flow rate of 1.0 ml min⁻¹ was used for the 12 min analysis. Retention times (min) were 5.48, 7.76, 8.69 and 9.60, and the limits of detection (ng) at 10 times background absorbance were 11.6, 6.8, 5.5 and 4.5 for *p,p'*-DDD, *p,p'*-DDT, *o,p'*-DDT and *p,p'*-DDE, respectively. None of the DDT-related chemicals were unstable in this assay. The data shown are representative of at least three independent experiments.

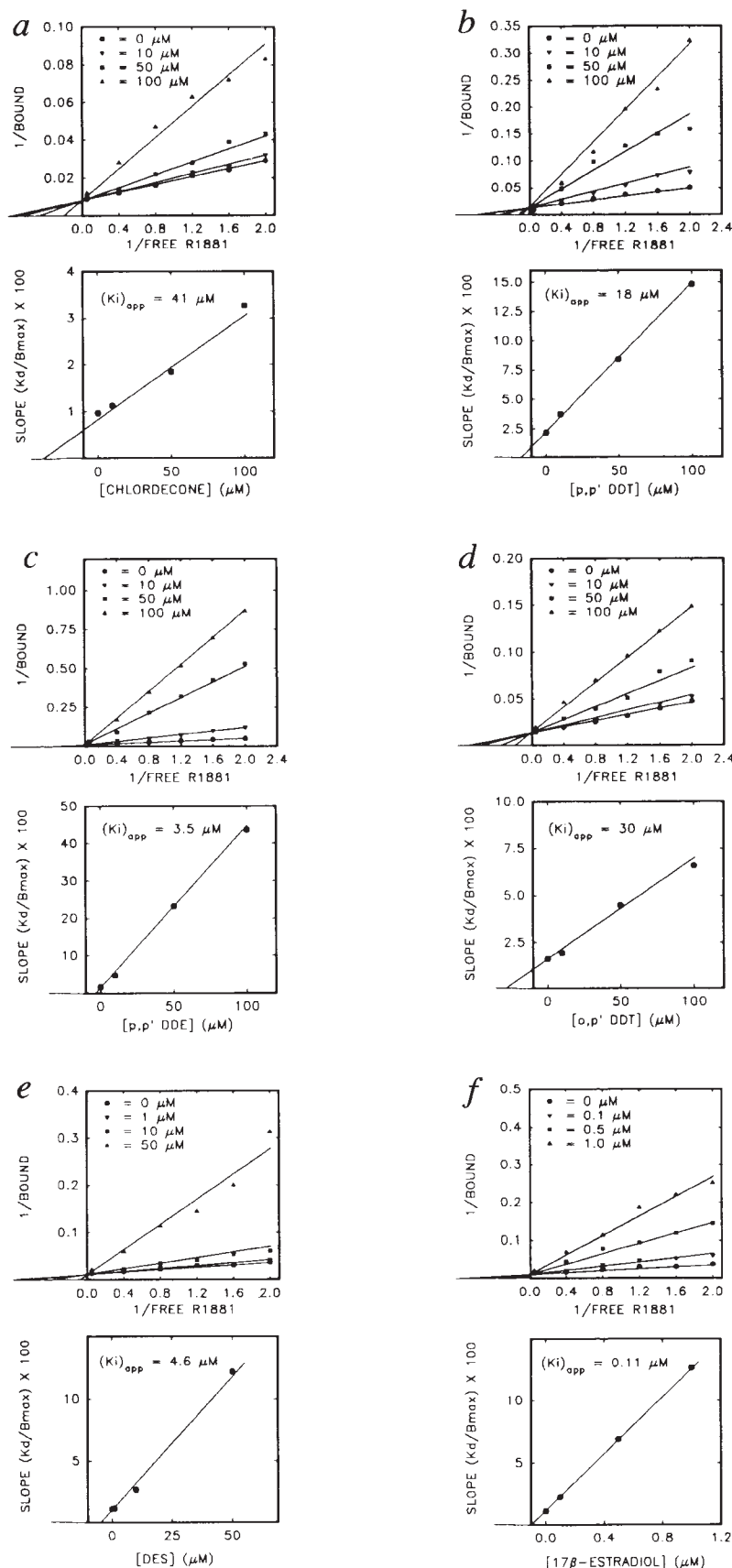
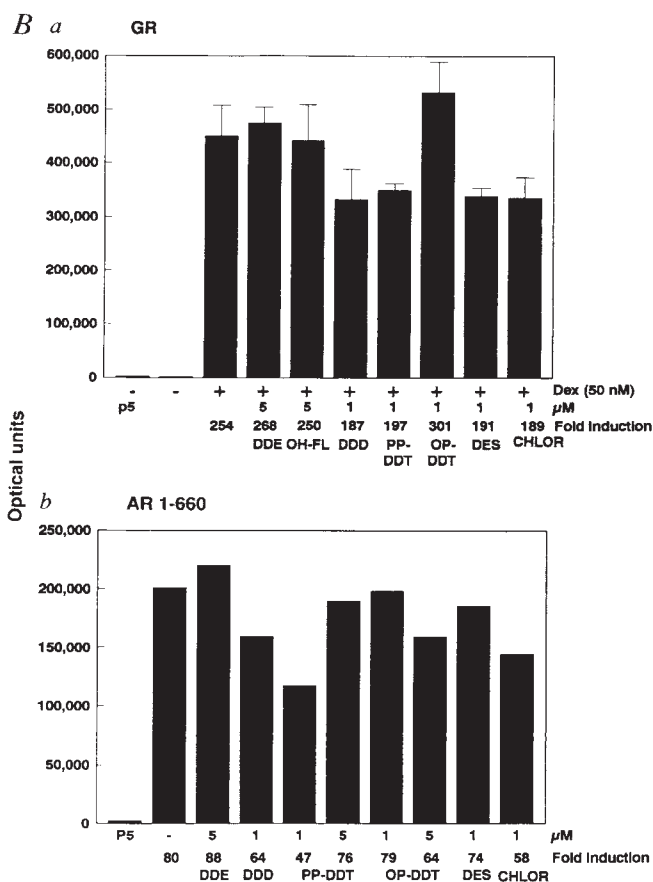
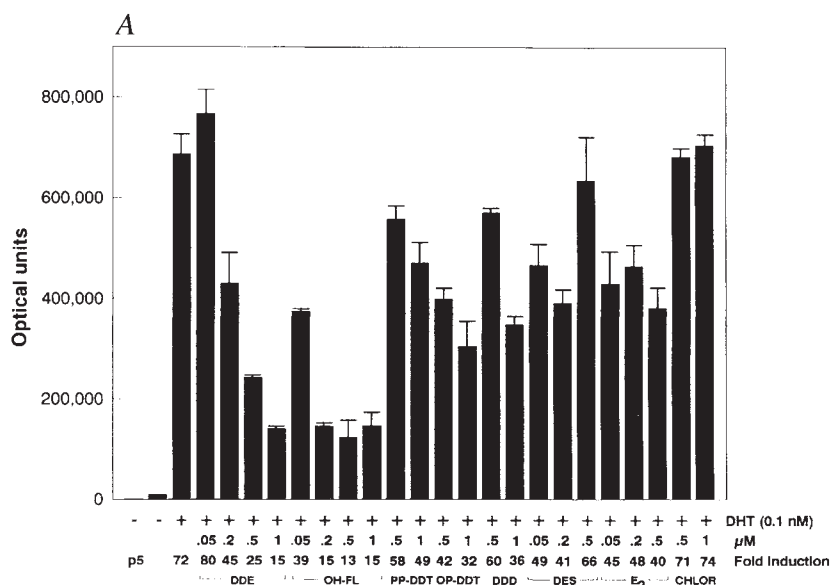


FIG. 3. A, Inhibition of AR transcriptional activation by 17β -oestradiol (E_2), DES and various suspected environmental endocrine disruptors including DDT isomers and metabolites. Transcriptional activity was assessed by transient cotransfection of 0.45×10^6 monkey kidney CV-1 cells per 6-cm dish with $0.1 \mu\text{g}$ pCMVhAR expression vector and $5 \mu\text{g}$ MMTV luciferase reporter vector (provided by R. M. Evans, Salk Institute, La Jolla, CA) using the calcium phosphate precipitation method³⁰. Twenty-four and 48 h after transfection, the indicated concentrations of hormones were added and 5 h after the last addition, cells were collected in 0.6 ml lysis buffer (Ligand Pharmaceuticals). Relative light units were determined using a Monolight 2010 Analytical Luminescence Laboratory luminometer. Indicated are the optical readings with standard error and fold induction relative to the activity in the absence of added hormone and are representative experiments of at least four independent determinations. None of the DDT-related chemicals were unstable or metabolized by the CV-1 cells during a 24 h period at 37°C as determined by HPLC/UV diode array spectroscopy following methylene chloride extraction of the assay media (see legend to Fig. 2). B, Lack of cytotoxic effects on glucocorticoid receptor (GR) transcriptional activation by various suspected environmental endocrine disruptors including DDT isomers and metabolites. Transcriptional activity was assessed by transient cotransfection of 0.45×10^6 monkey kidney CV-1 cells per 6-cm dish with either $0.1 \mu\text{g}$ of GR expression vector or the constitutively active expression vector AR1-660³⁰, lacking the steroid binding domain, together with $5 \mu\text{g}$ MMTV luciferase reporter vector using the calcium phosphate precipitation method as described above. Indicated are the optical readings with standard error and fold induction relative to the activity in the absence of added hormone and are representative of at least four independent determinations. a, Lack of inhibitory effects of various DDT isomers and metabolites on dexamethasone-induced human GR transcriptional activity. b, Lack of inhibitory effects of various DDT isomers and metabolites on constitutive AR1-660 transcriptional activity.



to increase steroid metabolism by increased steroid hydroxylase activity²¹.

The antiandrogenic effects of *p,p'*-DDE were also evident in 120-day-old adult male rats. Treatment with *p,p'*-DDE (200 mg per kg per day) by gavage for 4 days significantly reduced androgen-dependent seminal vesicle ($P < 0.01$) and ventral prostate ($P < 0.01$) weights relative to control rats, despite high serum testosterone levels (Fig. 1C). The ventral prostates from these *p,p'*-DDE-treated adult rats exhibited a 13-fold increase in androgen-repressed testosterone-repressed prostatic message 2 (TRPM-2) messenger RNA levels and a 35 per cent decline in androgen-induced prostate binding protein subunit 3 (C3) mRNA levels, relative to control rats²². These specific cellular responses are characteristic of the effects of antiandrogens on androgen-dependent gene expression in the rat ventral prostate^{23,24}. Taken together, these results provide compelling evidence that *p,p'*-DDE acts as an AR antagonist *in vitro* and *in vivo*.

The results support the hypothesis that the antagonistic effects of DDT on the male reproductive system are mediated by *p,p'*-DDE through inhibition of AR-androgen binding and subsequent inhibition of transcriptional activity. The concentration of *p,p'*-DDE required to inhibit AR transcriptional activity in cell culture ($0.2 \mu\text{M}$ or 63.6 parts per billion; p.p.b.) is less than levels that accumulate from the environment such as in the eggs of demasculinized male alligators in Florida's Lake Apopka ($5,800 \text{ p.p.b.}$)²⁵ and in humans in areas where DDT remains in use or is present in contaminated ecosystems. For example,

serum from individuals living in DDT-treated dwellings (such as for malaria control) in South Africa contain median DDT/DDE levels of 140 p.p.b. ²⁶. Moreover, in the mid-1960s, when DDT was still used in the United States, high concentrations of *p,p'*-DDE were found in tissues from stillborn infants in Atlanta, Georgia (650 p.p.b. in brain; 850 p.p.b. in lungs; $2,740 \text{ p.p.b.}$ in heart; 980 p.p.b. in liver; $3,570 \text{ p.p.b.}$ in kidney and 860 p.p.b. in spleen)²⁷. For comparison, DDT/DDE levels in the serum of pregnant rats given 25–200 mg DDT per kg per day in the diet

throughout gestation ranged from 700 to 1,700 p.p.b., whereas DDT/DDE levels in whole newborn rat pups averaged 2,100 p.p.b.²⁸ The above reports suggest that *p,p'*-DDE crosses the placenta to the developing human fetus and can reach levels known to inhibit human AR transcriptional activation *in vitro* and to induce antiandrogenic effects in rats *in vivo*. Taken together, these results suggest that the reported increased incidence of developmental male reproductive system abnormalities in wildlife and humans may reflect antiandrogenic activity of the persistent DDT metabolite *p,p'*-DDE, acting at the level of the AR. □

Received 3 April; accepted 25 April 1995.

1. Sharpe, R. M. & Skakkebaek, N. E. *Lancet* **341**, 1392–1395 (1993).
2. Carlson, E., Giwercman, A., Keiding, N. & Skakkebaek, N. *Br. med. J.* **305**, 609–612 (1992).
3. Nelson, J. A. *Biochem. Pharmacol.* **23**, 447–451 (1974).
4. Robison, A. K., Schmidt, W. A. & Stancel, G. M. *J. Tox. envir. Hlth* **16**, 493–508 (1985).
5. Welch, R. M., Levin, W. & Conney, A. H. *Tox. appl. Pharmac.* **14**, 358–367 (1969).
6. Gellert, R. J., Heinrichs, W. L. & Swerdloff, R. S. *Endocrinology* **91**, 1095–1100 (1972).
7. Clemens, L. G., Gladue, B. A. & Coniglio, L. P. *Horm. Behav.* **10**, 40–53 (1978).
8. Neumann, F. et al. *Recent Prog. Horm. Res.* **26**, 337–410 (1970).
9. Wilson, E. M. & French, F. S. *J. biol. Chem.* **251**, 5620–5629 (1976).

10. Kelce, W. R., Monosson, E., Gamcsik, M. P., Laws, S. C. & Gray, L. E. *Tox. appl. Pharmac.* **126**, 276–285 (1994).
11. Morgan, D. P. & Roan, C. C. *Archs. envir. Hlth* **22**, 301–308 (1971).
12. Dustman, E. H. & Stickel, L. F. *Ann. N.Y. Acad. Sci.* **160**, 162–172 (1969).
13. Bitman, J. & Cecil, H. C. *J. agric. Fd Chem.* **18**, 1108–1112 (1970).
14. Nelson, J. A. *Biochem. Pharmacol.* **23**, 447–451 (1974).
15. Nelson, J. A., Struck, R. F. & James, R. J. *Tox. envir. Hlth* **4**, 325–339 (1978).
16. Zhou, Z.-X., Sar, M., Simental, J. A., Lane, M. V. & Wilson, E. M. *J. biol. Chem.* **269**, 13115–13123 (1994).
17. Imperato-McGinley, J., Sanchez, R. S., Spencer, J. R., Yee, B. & Vaughan, E. D. *Endocrinology* **131**, 1149–1156 (1992).
18. Dürnberger, H. & Kratochwil, K. *Cell* **19**, 465–471 (1980).
19. Korenbrot, C. C., Huhtaniemi, I. T. & Weiner, R. I. *Biol. Reprod.* **17**, 298–303 (1977).
20. Gray, L. E., Ostby, J. M. & Kelce, W. R. *Tox. appl. Pharmac.* **129**, 46–52 (1994).
21. Kupfer, D. *CRC Crit. Rev. Tox.* **4**, 83–124 (1975).
22. Stone, C. R., Gray, L. E. Jr, Roberts, K. P. & Kelce, W. R. *J. Androl.* **16** (suppl.), 42 (1995).
23. Leger, J. G., Guellecand, R. L. & Tenniswood, M. P. R. *The Prostate* **13**, 131–142 (1988).
24. Page, M. J. & Parker, M. G. *Molec. cell. Endocrin.* **27**, 343–355 (1982).
25. Guillette, L. J. Jr et al. *Envir. Hlth Perspect.* **102**, 680–688 (1994).
26. Bouwman, H., Cooppan, R. M., Becker, P. J. & Ngxongo, S. J. *Tox. Envir. Hlth* **33**, 141–155 (1991).
27. Curley, A., Copeland, M. F. & Kimbrough, R. D. *Archs. envir. Hlth* **19**, 628–632 (1969).
28. Cobb, J., Ostby, D., Huey, G. M. & Talens, G. M. *Tox. appl. Pharmac.* **18**, 907–916 (1971).
29. Kelce, W. R. & Ganjam, V. K. *Steroids* **52**, 217–235 (1988).
30. Simental, J. A., Sar, M., Lane, M. V., French, F. S. & Wilson, E. M. *J. biol. Chem.* **266**, 510–518 (1991).

ACKNOWLEDGEMENTS. We thank F. S. French and G. R. Klinefelter for reviewing the manuscript, M. Cobb, J. Ostby, D. Huey and S. Carey for technical assistance and M. R. Hollowbush for preparation of several figures. Supported by grants from the National Institute of Child Health and Human Development Center for Population Research.

Elevation of auditory thresholds by spontaneous cochlear oscillations

Nicholas L. Powers, Richard J. Salvi*, Jian Wang, Vlasta Spongr & Chun Xiao Qiu

Hearing Research Laboratory, Department of Communicative Disorders & Sciences, 215 Parker Hall, State University of New York at Buffalo, Buffalo, New York 14214, USA

THE inner ear sometimes acts as a robust sound generator, continuously broadcasting sounds (spontaneous otoacoustic emissions) which can be intense enough to be heard by other individuals standing nearby^{1–4}. Paradoxically, most individuals are unaware of the sounds generated within their ears. Two hypotheses could explain this paradox: (1) the spontaneous emissions may not be transmitted to the central nervous system; or (2) the spontaneous emission produces a continuous, high rate of neural activity, which, like the natural pattern of spontaneous activity, is ignored by the central nervous system. Here we demonstrate that high-intensity spontaneous otoacoustic emissions can vigorously activate auditory nerve fibres in mammals (*Chinchilla laniger*). This 'internal biological noise' creates a 'line busy' signal that significantly degrades a neuron's ability to respond to sound and results in a hearing loss completely different from that caused by damage to sensory cells^{1,4}.

Robust spontaneous otoacoustic emissions between 4 and 5 kHz were detected in the ear canal of two chinchillas after exposure to intense sounds (e.g. Fig. 1a). The spontaneous emissions could be suppressed by external tones within a narrow frequency range (Fig. 1b), suggesting that they originated from a narrow region of the cochlea. Distortion product otoacoustic emissions ($2F_1-F_2$) elicited by two external tones were normal at frequencies near the spontaneous emission (e.g. Fig. 1c), implying that the cochleae and sensory outer hair cells were functionally intact^{7,8}. Spontaneous otoacoustic emissions presumably arise from electromotile oscillations in outer hair cells^{9,10}. Consistent with this view, salicylates, which can transiently abolish the electromotility of outer hair cells¹¹, temporarily eliminated one chinchilla's spontaneous otoacoustic emission¹².

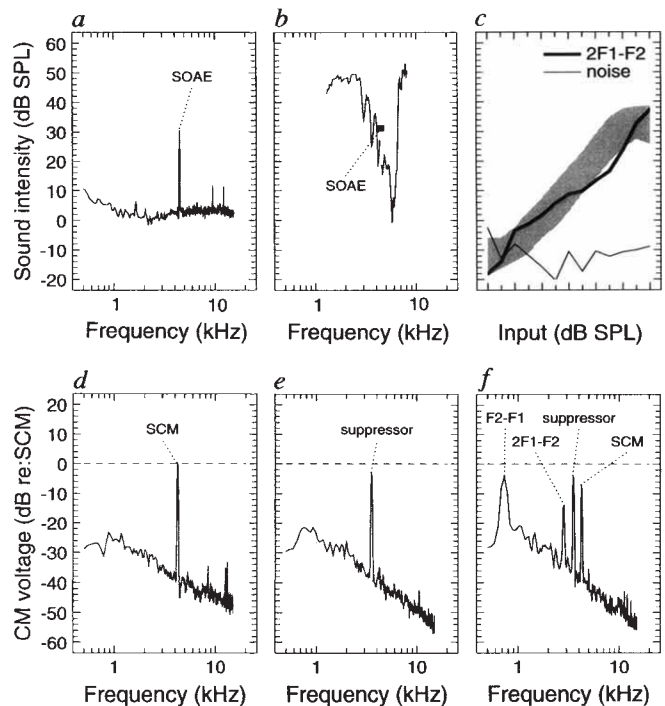
Outer hair cells constitute 75% of the sensory cell population, and are the predominant generators of the cochlear microphonic, a gross potential that mirrors the stimulus waveform¹³. In quiet conditions a high-voltage, spontaneous cochlear microphonic was observed in both animals¹⁴. The spectrum of the spontaneous cochlear microphonic shown in Fig. 1d contained one large peak (4147 Hz) at the same frequency as the spontaneous otoacoustic emission. An external tone of 3481 Hz and 61 dB sound pressure level (SPL) completely suppressed this spontaneous cochlear microphonic, and produced a stimulus-evoked cochlear microphonic (Fig. 1e) that had an amplitude equal to that of the spontaneous cochlear microphonic (Fig. 1d). This indicates that the amplitude of the spontaneous mechanical vibration within the cochlea is equivalent to that produced by a 61 dB SPL external tone; this suggests that the spontaneous otoacoustic emission is attenuated by approximately 30 dB when it is transmitted from the cochlea to the ear canal¹⁵. When the level of the suppressor tone was reduced to 54 dB SPL, the spontaneous cochlear microphonic reappeared (Fig. 1f), along with two additional distortion tones, $2F_1-F_2$ and F_2-F_1 . This shows that the external tone interacts with the spontaneous mechanical vibration and results in distortion products in the cochlear microphonic which are similar to acoustic distortion products measured in the ear canal.

Although inner hair cells comprise 25% of all sensory cells, they synapse, one to one, on 90 to 95% of the auditory nerve fibres that project centrally¹⁶. To determine if the spontaneous emission affected auditory nerve sensitivity, compound action potential thresholds were measured (Figs 2a and 3c). Thresholds were elevated by 40 dB near the emission frequency (Figs 1a and 3a), suggesting that the 4 kHz region of the cochlea was damaged^{1,4,5}. However, this interpretation conflicts with histological results showing intact hair cells in the 4.2-kHz region of the cochlea (Fig. 2b). Apparently, hair-cell lesions are not a necessary condition for producing spontaneous emissions¹⁷.

Recordings were obtained from single fibres of the auditory nerve¹⁸. Because each neuron's characteristic frequency (CF) is precisely related to its position in the cochlea, the activity of each neuron^{19,20} provides detailed information about the functional status from a narrow segment of the cochlea. Neurons with CFs near the frequency of the spontaneous emission had thresholds that were elevated by approximately 40 dB (Figs 2a and 3c); however, their narrow frequency-threshold tuning curves implied normal cochlear function^{18,20,21} (Figs 2d and 3d).

* To whom correspondence should be addressed.

FIG. 1 Otoacoustic emissions measured with a microphone (dB SPL re 20 μ Pa; Etymotic ER10B) in ear canal (a–c) and cochlear microphonic (CM) potential (d–f) measured from round window of cochlea (chinchilla 1). a, Average ($n=32$) spectrum of spontaneous otoacoustic emission (SOAE) in ear canal²⁸. b, Frequency–intensity combinations of external tone that suppressed the SOAE (square) by 3 dB. c, Amplitude of cubic distortion product otoacoustic emission, $2F_1-F_2$ (thick line, 3200 Hz), when two primary tones, F_1 (4000 Hz) and F_2 (4800 Hz), of equal sound pressure level (SPL), were varied from 10 to 70 dB SPL²⁸. Background noise (thin line) in ear canal at $F_n=0.7(2F_1-F_2)$. Shaded area: mean (± 1 s.d.) distortion product amplitude in normal chinchilla ($n=47$). d, Average ($n=32$) spectrum of CM voltage in quiet conditions¹⁸. Peak in spectrum near 4.2 kHz is spontaneous cochlear microphonic (SCM); 0 dB equals 39 μ V r.m.s. e, Average ($n=32$) amplitude spectrum of CM when a 61 dB SPL, 3481 Hz suppressor tone was presented. Note peak in CM voltage at 3481 Hz and complete suppression of SCM. CM amplitude at 3481 Hz is approximately equal to SCM in d. f, Average ($n=32$) spectrum of CM when 3481 Hz tone was presented at 54 dB SPL. Peak in CM at 3481 Hz corresponds to external suppressor tone. External suppressor tone partially suppressed (-7 dB) the SCM near 4.2 kHz. External suppressor tone and SCM interact to produce distortion products in CM at $2F_1-F_2$ and F_2-F_1 . Chinchilla 1 was exposed to pure tone approximately 2 years earlier (2.8 kHz, 2 h at 105 dB SPL).



Neurons with higher and lower CFs had normal thresholds and frequency selectivity.

Auditory neurons have spontaneous discharge rates that normally vary from 0 to 100 spikes per second²². Significantly, the spontaneous discharge rates in the two chinchillas with spontaneous emissions were higher than normal (Kruskal-Wallis, one-way ANOVA on ranks, $P=0.0018$). In mammals, spontaneous neural activity cannot be suppressed by a single tone^{23,24}. However, neural activity activated by one tone at CF can be suppressed by a second tone at a frequency higher or

lower than CF. This nonlinear, two-tone suppression effect is present in healthy ears but absent in damaged ears^{18,21,23}. If the high 'spontaneous' discharge rates observed in neurons with CFs near the spontaneous emission were due to an internally generated tone, then it should be possible to demonstrate two-tone suppression using only a single external tone^{18,21,23}. As predicted, spontaneous activity could be suppressed with a single tone, but only in neurons with CF near the frequency of the spontaneous emission (Figs 2c, d and 3d). The latency of suppression was nearly zero²⁴, thereby ruling out an inhibitory synapse and

FIG. 2 Physiological and anatomical data from chinchilla 1. a, Compound action potential (CAP) recordings obtained from round window electrode; single auditory nerve fibre recording made with glass micropipettes^{18,22}. Mean CAP threshold (± 1 s.d.) in normal control animals (shaded area; $n=17$). CAP threshold from animal with spontaneous otoacoustic emission (thin line). Approximate minimum thresholds versus CF in normal auditory nerve fibres ($n=233$; thick line)²². Threshold versus CF in auditory nerve fibres ($n=100$) from animal with spontaneous emission. b, Cytocochleogram shows percentage of missing inner hair cells (IHC) and outer hair cells (OHC) versus position in cochlea. Cochlea was embedded in Spurr resin, prepared as a flat surface preparation and examined by differential interference contrast microscopy¹⁸. Position in the cochlea was transformed to frequency using frequency-place map²⁹. Location of spontaneous otoacoustic emission (SOAE) is indicated. c, Poststimulus time histogram showing number of spike discharges in relation to suppressor tone (50 ms, 1 ms rise/fall time, 1294 Hz, 90 dB SPL). Recording from auditory nerve fibre with CF (5176 Hz; threshold 58 dB SPL, spontaneous rate 155 spikes per s) near spontaneous emission. Plateau of suppressor tone occurs 12.5 ms from the start of the poststimulus time histogram. Latency of neuron's response includes transmission time delays through middle ear, cochlea (0.7 ms), hair cell–nerve fibre synapse (0.7 ms) and axonal conduction time. Spontaneous activity was completely suppressed by tone. d, Excitatory tuning curve (thick line) and lower boundary of single-tone suppression area (thin line) from auditory nerve fibre (CF 4746 Hz, threshold 48 dB SPL, spontaneous rate 116 spikes per s) in chinchilla with spontaneous otoacoustic emission. Note narrow tip of excitatory tuning curve. Excitatory tuning curve and single-tone suppression area measured with computer-automated threshold-tracking procedure^{18,22} (tone on, 50 ms; tone off, 50 ms; threshold criterion, 1 spike difference between tone-on and tone-off intervals).

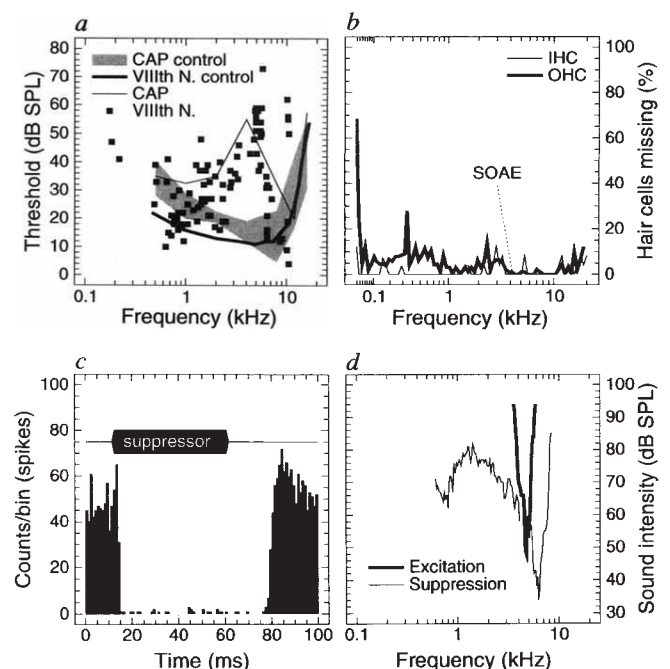
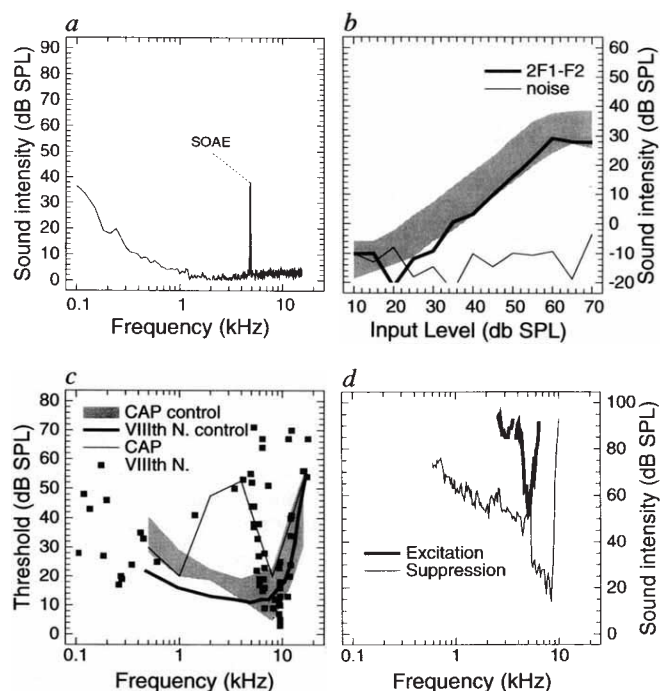


FIG. 3 Results from chinchilla 2. *a*, SOAE measured with a microphone in the ear canal (see Fig. 1*a*)²⁸. *b*, amplitude of cubic distortion product otoacoustic emission, $2F_1-F_2$ (thick line, 3200 Hz), when two primary tones, F_1 (4000 Hz) and F_2 (4800 Hz), of equal SPL, were varied from 10 to 70 dB SPL²⁸. Background noise (thin line) in ear canal at $F_n = 0.7(2F_1-F_2)$. Shaded area indicates mean (± 1 s.d.) distortion product amplitude in normal chinchilla ($n=47$). *c*, Compound action potential (CAP) recordings obtained from round window electrode; single auditory nerve fibre recording made with glass micropipettes^{18,22}. Mean CAP threshold (± 1 s.d.) in normal animals (shaded area; $n=17$). CAP threshold from animal with spontaneous otoacoustic emission (thin line). Approximate minimum threshold versus CF in normal auditory nerve fibres ($n=233$; thick line)²². Threshold versus CF in auditory nerve fibres ($n=69$) from animal with spontaneous emission. *d*, Excitatory tuning curve (thick line) and lower boundary of single-tone suppression area (thin line) from auditory nerve fibre (CF 5065 Hz, threshold 52 dB SPL, spontaneous rate 181 spikes per s) in chinchilla with spontaneous otoacoustic emission. Chinchilla 2 was exposed to impact noise approximately 6 months earlier (200 ms B duration, 1 per s, 125 dB peak SPL).



implicating a mechanical interaction between the external tone and internal oscillator^{4,7}. Neural suppression persisted approximately 10 ms longer than the suppressing tone, which is consistent with spontaneous emission suppression data⁵. Single-tone suppression boundaries in animals with spontaneous emissions (Figs 2*d* and 3*d*) were similar to two-tone suppression boundaries observed in normal animals^{21,23,24}. However, the threshold for single-tone suppression above CF was often lower than the excitatory threshold consistent with the acoustic suppression hypersensitivity shown in Fig. 1*b*.

The low-level spontaneous emissions that occur naturally in some ears do not appear to influence threshold³⁰, whereas the less common, high-level spontaneous emissions reported here elevated neural thresholds significantly. Normal tuning, distortion product emissions (Figs 1*c* and 3*b*) and hair cells were

observed near the emission frequency, yet neural thresholds were much higher than normal^{1,4}. Significantly, 'single tone' suppression was observed in high spontaneous rate neurons with CFs near the emission frequency, and the suppression boundaries resembled the two-tone suppression boundaries seen in normal mammals. We conclude that excitatory thresholds are elevated because spontaneous cochlear oscillations induce high rates of 'spontaneous activity' in neurons with CFs near the emission frequency. The high 'spontaneous rates' create a 'line busy' signal that significantly increases neural thresholds to sounds^{25,26}, resulting in a hearing loss that is fundamentally different from that caused by hair-cell damage^{20,21,27}. Apparently, some humans with intense spontaneous emissions owe their hearing loss to internal 'noise' which they are unable to perceive^{1,2}. □

Received 28 December 1994; accepted 19 April 1995.

1. Wilson, J. P. & Sutton, G. J. in *Tinnitus* (eds Evered, D. & Lawrenson, G.) 82–97 (Pitman, London, 1981).
2. Kemp, D. T. *J. acoust. Soc. Am.* **64**, 1386–1391 (1978).
3. Zurek, P. *J. acoust. Soc. Am.* **69**, 514–523 (1981).
4. Ruggero, M. A., Kramak, B. & Rich, N. C. *Hear. Res.* **13**, 293–296 (1984).
5. Zurek, P. M. & Clark, W. W. *J. acoust. Soc. Am.* **70**, 446–450 (1981).
6. Wilson, J. P. *Hear. Res.* **2**, 233–252 (1980).
7. Zurek, P. M. & Clark, W. W. *J. acoust. Soc. Am.* **72**, 774–780 (1982).
8. Subramaniam, M., Salvi, R. J., Spongr, V. P., Henderson, D. & Powers, N. L. *Hear. Res.* **74**, 204–216 (1994).
9. Brownell, W. E., Bader, C. R., Bertrand, D. & de Ribaupierre, Y. *Science* **227**, 194–196 (1985).
10. Ashmore, J. F. *J. Physiol., Lond.* **388**, 323–347 (1987).
11. Dieler, R., Shehata-Dieler, W. & Brownell, W. J. *Neurocyt.* **20**, 637–653 (1991).
12. McFadden, D. & Plattsmier, H. S. *J. acoust. Soc. Am.* **76**, 443–448 (1984).
13. Dallos, P. & Cheatham, M. A. *J. acoust. Soc. Am.* **60**, 510–512 (1976).
14. Evans, E. F., Wilson, J. P. & Borenstein, T. A. in *Tinnitus* (eds Evered, D. & Lawrenson, G.) 108–139 (Pitman, London, 1981).
15. Nedzelitsky, V. *J. acoust. Soc. Am.* **66**, 1676–1689 (1980).
16. Spoendlin, H. *Acta otol.* **73**, 235–248 (1972).
17. Clark, W. W., Kim, D. O., Zurek, P. M. & Bohne, B. A. *Hear. Res.* **16**, 299–314 (1984).

18. Salvi, R. J., Perry, J., Hamernik, R. P. & Henderson, D. in *New Perspectives on Noise-Induced Hearing Loss* (eds Hamernik, R. P., Henderson, D. & Salvi, R. J.) 165–188 (Raven, New York, 1982).
19. Liberman, M. C. *Science* **217**, 1239–1241 (1982).
20. Liberman, M. C., Dodds, L. W. & Learson, D. A. in *Basic and Applied Aspects of Noise-Induced Hearing Loss* (eds Salvi, R. J., Henderson, D., Hamernik, R. P. & Colletti, V.) 163–177 (Plenum, New York, 1986).
21. Schmiedt, R. A., Zwislocki, J. J. & Hamernik, R. P. *J. Neurophysiol.* **43**, 1367–1389 (1980).
22. Powers, N. thesis, SUNY at Buffalo (1991).
23. Sachs, M. B. & Kiang, N. Y.-S. *J. acoust. Soc. Am.* **43**, 1120–1128 (1968).
24. Kiang, N. Y.-S., Watanabe, E. C., Thomas, L. F. & Clark, L. F. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, 1965).
25. Sinex, D. G. & Havey, D. C. *Hear. Res.* **13**, 285–292 (1984).
26. Sinex, D. G. & Havey, D. C. *J. Neurosci.* **56**, 1763–1780 (1986).
27. Kiang, N. Y.-S., Moxon, E. C. & Levine, R. A. in *Sensorineural Hearing Loss* (ed. Wolstenholme G. E. W. & Knight, J.) 241–272 (Churchill, London, 1970).
28. Froyovich, O., Rebala, V., Salvi, R. J. & Rassael, H. *J. acoust. Soc. Am.* **97**, 3021–3029 (1995).
29. Eldredge, D., Miller, J. D. & Bohne, B. A. *J. acoust. Soc. Am.* **69**, 1091–1095 (1981).
30. Lonsbury-Martin, B. L., Harris, F., Stagner, B. B., Hawkins, M. D. & Martin, G. K. *Ann. Otol. Rhinol. Lar.* **99**, 15–29 (1990).

ACKNOWLEDGEMENTS. This work was supported by the NIDCD.

Mediation of PACAP-like neuropeptide transmission by coactivation of Ras/Raf and cAMP signal transduction pathways in *Drosophila*

Yi Zhong

Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

MUCH work on the signal transduction mechanisms underlying neurotransmission has been directed towards studying the roles of the cyclic AMP and phosphoinositide pathways¹⁻³. Upon ligand binding, the transmitter receptors interact with heterotrimeric G proteins, allowing G_α and $G_{\beta\gamma}$ subunits to disengage^{2,3}. The free G_α then modulates the activity of adenylyl cyclase and phospholipase C¹⁻³. It has been suggested that the $G_{\beta\gamma}$ complex which is activated through muscarinic or neuropeptide receptors can stimulate mitogen-activated protein kinase (MAPK) via activation of the small guanine-nucleotide-binding protein Ras^{4,5}. Sequential activation of the intermediates in the Ras/Raf serine-threonine protein kinase/MAPK kinase/MAPK/transcription factor pathway has emerged as a central mechanism for controlling cell proliferation and differentiation in yeast, worms, fruitflies and mammals⁶⁻¹¹. Here we show, by analysis of *Drosophila* mutants, that synaptic current and modulation of K^+ current, triggered by a pituitary adenylyl cyclase-activating polypeptide-like neuropeptide¹², are mediated by coactivation of the Ras/Raf and Rutabaga-adenylyl cyclase pathways. Thus the Ras/Raf pathway also appears to be essential for G-protein-coupled neurotransmission.

Western blot, immunohistochemical and electrophysiological analyses have shown that a pituitary adenylyl cyclase-activating polypeptide¹³ (PACAP38)-like neuropeptide functions as a neurotransmitter at the body-wall neuromuscular junction of

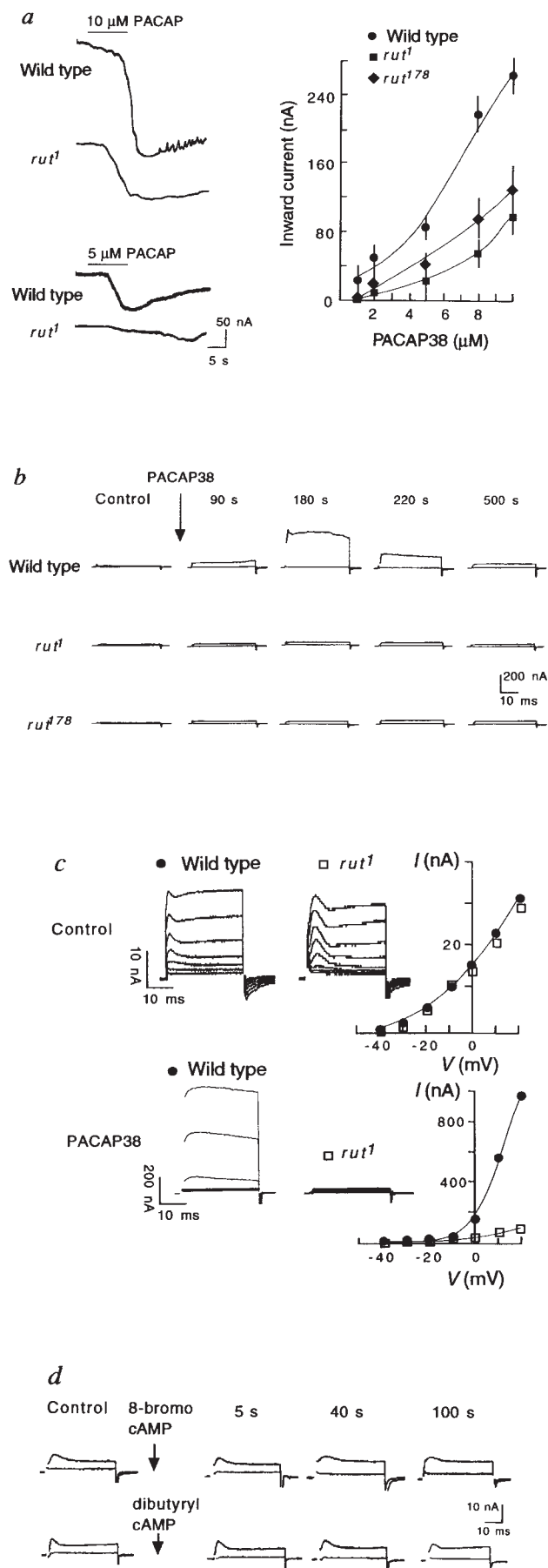


FIG. 1 Diminished PACAP38 response in *rut* mutants. **a**, Reduced PACAP38-induced synaptic inward current in *rut¹* and *rut¹⁷⁸* mutants. Left, synaptic current traces; the horizontal bar above them indicates the period during which PACAP38 was perfused. Right, PACAP38 dose dependence of the peak inward current (mean $\pm$ s.e.m.) among wild-type, *rut¹* and *rut¹⁷⁸* mutant larvae; *n* (number of muscle fibres recorded) = 26 (4 to 6 for each concentration), 18 (3 to 4 for each concentration), 17 (3 to 4 for each concentration) for wild type, *rut¹* and *rut¹⁷⁸*, respectively. **b**, PACAP38-induced enhancement of voltage-activated K^+ currents is abolished in *rut* mutants. The arrow indicates focal application of PACAP38 (5 μ M). In this and all following figures, the time in seconds after pressure-ejection of PACAP38 is indicated on the top of current traces. **c**, Current-voltage (*I*-*V*) relations before or 4 min after application of PACAP38 (5 μ M). The *I*-*V* curve demonstrates the voltage dependence of the modulation. The data presented in **b** and **c** are chosen from samples *n* = 10, 12 and 9 for wild type *rut¹* and *rut¹⁷⁸*, respectively. All of the wild-type but none of mutant muscle fibres responded. **d**, 8-bromo cAMP and dibutyryl cAMP (1 mM and 10 mM) fail to mimic PACAP38 response in wild-type larvae, although K^+ currents were enhanced slightly²⁸; *n* = 5, 4. Data in this and all following figures, except those shown in Fig. 4, were recorded with the bath saline containing 1 mM Ca^{2+} .

METHODS. The larval body-wall neuromuscular preparation has been described previously^{12,28-30}. The set-up, saline²⁹, recording conditions and voltage paradigms are as described previously¹². For recording the PACAP38-induced synaptic current, the membrane potential was clamped at -80 mV. For recording K^+ currents, command voltages were stepped from the holding potential of -80 to -50 and 0 mV, respectively. These currents include outward K^+ and inward Ca^{2+} currents, but the inward Ca^{2+} component is completely masked²⁸.

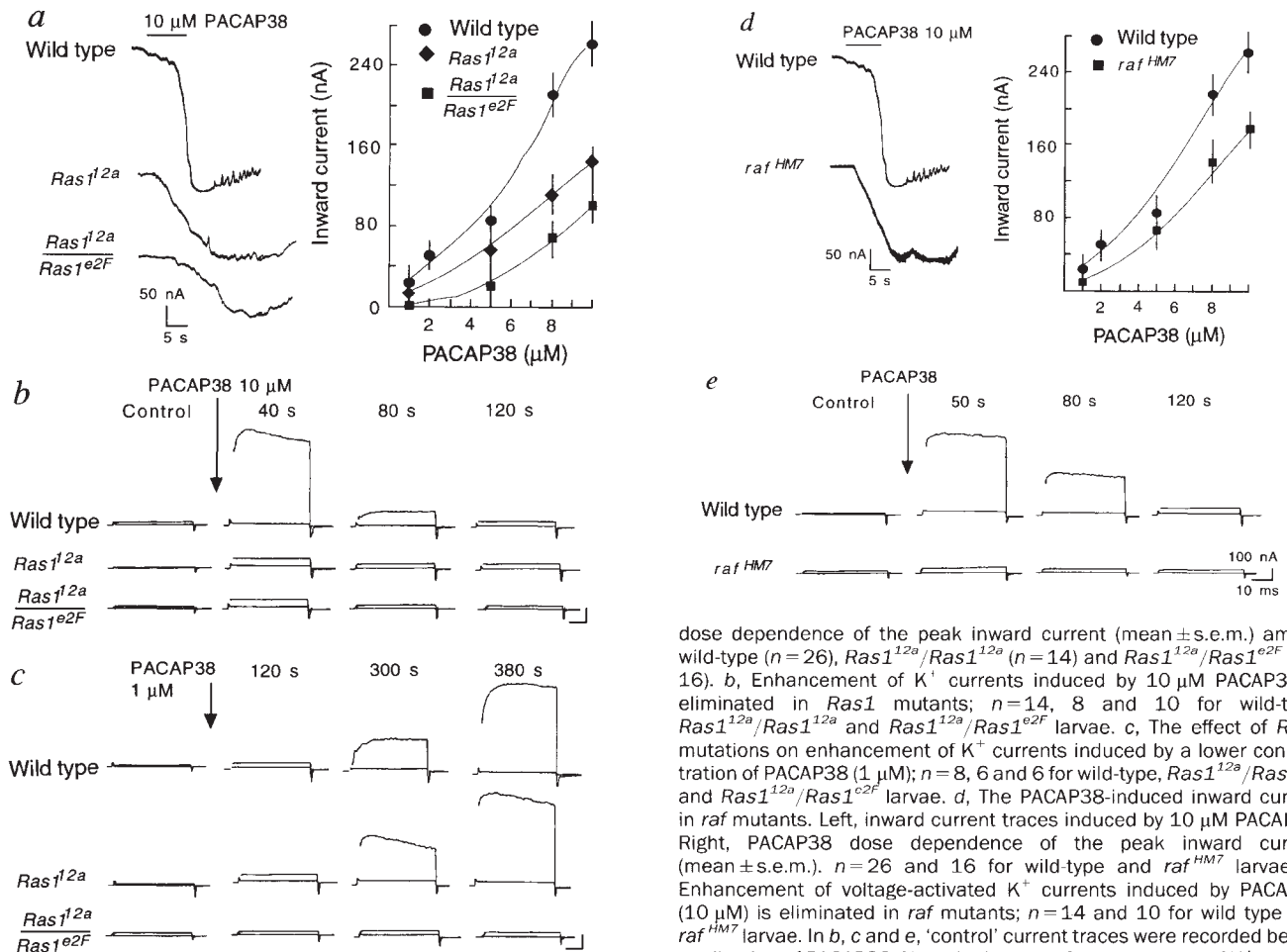


FIG. 2 Diminished PACAP38 response in *Ras1* and *raf* mutants. **a**, The PACAP38-induced inward current is reduced in *Ras1* mutants. Left, inward current traces induced by 10 μ M PACAP38. Right, the PACAP38

dose dependence of the peak inward current (mean $\pm$ s.e.m.) among wild-type ($n=26$), *Ras1*^{12a}/*Ras1*^{12a} ($n=14$) and *Ras1*^{12a}/*Ras1*^{e2F} ($n=16$). **b**, Enhancement of K⁺ currents induced by 10 μ M PACAP38 is eliminated in *Ras1* mutants; $n=14$, 8 and 10 for wild-type, *Ras1*^{12a}/*Ras1*^{12a} and *Ras1*^{12a}/*Ras1*^{e2F} larvae. **c**, The effect of *Ras1* mutations on enhancement of K⁺ currents induced by a lower concentration of PACAP38 (1 μ M); $n=8$, 6 and 6 for wild-type, *Ras1*^{12a}/*Ras1*^{12a} and *Ras1*^{12a}/*Ras1*^{e2F} larvae. **d**, The PACAP38-induced inward current in *raf* mutants. Left, inward current traces induced by 10 μ M PACAP38. Right, PACAP38 dose dependence of the peak inward current (mean $\pm$ s.e.m.). $n=26$ and 16 for wild-type and *raf*^{HM7} larvae. **e**, Enhancement of voltage-activated K⁺ currents induced by PACAP38 (10 μ M) is eliminated in *raf* mutants; $n=14$ and 10 for wild type and *raf*^{HM7} larvae. In **b**, **c** and **e**, 'control' current traces were recorded before application of PACAP38. Note the latency of enhancement of K⁺ current is very different in response to 10 versus 1 μ M PACAP38. Also note that *Ras1*^{12a}/*Ras1*^{12a} mutants show normal enhancement of K⁺ currents in response to 1 μ M PACAP38 but no response to 10 μ M. Scale bars in **b** and **c** are horizontal, 10 ms; vertical, 100 nA.

Drosophila larvae¹². The peptide is released upon high-frequency-like stimulation of motor axons and induces a biphasic response: a slow synaptic inward current is followed by a 100-fold increase in voltage-activated K⁺ conductance¹². The synaptic current leads to depolarization of muscle membrane¹², and the enhancement of K⁺ currents is expected to retard the generation of, or shorten the duration of, the action potential¹⁴. This evoked response can be mimicked by application of vertebrate PACAP38 to wild-type muscle fibres (Fig. 1; ref. 12). The slow synaptic current was observed immediately after perfusion of PACAP38, while the enhancement of K⁺ currents was delayed; higher concentrations of PACAP38 yielded shorter latencies.

Because PACAP38 activates an adenylyl cyclase in vertebrates^{13,15}, the PACAP38 response was examined in the mutant *rutabaga* (*rut*), a gene encoding the calcium/calmodulin (Ca²⁺/CaM)-sensitive adenylyl cyclase^{16,17}. *Rutabaga*-adenylyl cyclase (*Rut*-AC) activity is completely eliminated in the *rut*¹ mutant^{16,17}, and its immunoreactivity is nearly abolished in the *rut*¹⁷⁸ mutant¹⁸. The PACAP38-induced synaptic current was significantly reduced but not eliminated in *rut*¹ and *rut*¹⁷⁸ mutants (Fig. 1a). In contrast, enhancement of K⁺ current was abolished in both mutants in response to 10 μ M (Fig. 1b, c), 5 μ M or 1 μ M (not shown) PACAP38. These results suggest that PACAP38 may activate *Rut*-AC to synthesize cAMP, which in turn leads to modulation of ion channels. Applications of the membrane-permeable cAMP analogues, 8-bromo cAMP, dibu-

tyryl cAMP (Fig. 1d) or Sp-8-Br-cAMP (not shown), however, elicited neither the PACAP38-like synaptic current nor the 100-fold enhancement of K⁺ currents. This observation suggested that coactivation of other signal transduction pathways might be required to produce a PACAP38-like response. To pursue this further, the *Ras*/*Raf* pathway¹⁹ was investigated.

Two *Ras1* mutant alleles were examined. *Ras1*^{12a} is a hypomorphic mutation which reduces normal levels of *Ras1* protein (C. Berg, personal communication). Flies homozygous for this allele are viable. *Ras1*^{e2F} is a more severe mutation; homozygous flies are lethal²⁰. Heteroallelic *Ras1*^{12a}/*Ras1*^{e2F} flies, however, are viable.

The effects of *Ras1* mutations on the PACAP38 response were very similar to those observed in *rut* mutants. The synaptic current was reduced (Fig. 2a), and the enhancement of K⁺ currents induced by 10 μ M PACAP38 was absent (Fig. 2b) in *Ras1*^{12a} and *Ras1*^{12a}/*Ras1*^{e2F} larvae. Surprisingly, if a lower concentration of PACAP38 (1 μ M) was perfused, enhancement of K⁺ currents appeared to be normal in homozygous *Ras1*^{12a} larvae (Fig. 2c; see Fig. 4c legend and last paragraph of the text). This enhancement did not occur in more severe *Ras1*^{12a}/*Ras1*^{e2F} mutants (Fig. 2c).

Ras proteins may activate several targets such as GTPase-activating proteins (GAPs)²¹, *Raf*²² or phosphatidylinositol-3-OH kinase²³. The *Ras*/GAP complex appears to inhibit coupling of muscarinic receptors to atrial K⁺ channels^{21,24}. Examination

of two mutant alleles of a *Drosophila* GAP gene (*r1533^{B1}* and *r7533^{PB}*)²⁵ revealed a normal PACAP38 response (not shown). However, recordings from a weakly viable *raf^{HM7}* allele²² revealed slightly reduced synaptic current and no enhancement of K⁺ currents in response to either 1 μ M (not shown) or 10 μ M PACAP38 (Fig. 2d, e). These observations, together with results from *Ras1* mutants, suggest that the Ras/Raf pathway, in addition to the cAMP cascade, may be necessary for the enhancement of K⁺ currents induced by exogenous PACAP38.

To test the endogenous neuropeptide response¹², a wild-type neuromuscular preparation was bathed in saline containing 1.5 mM Ca²⁺, and 40 Hz stimulation was delivered for 3 s. Under these conditions, a PACAP38-like enhancement of K⁺ currents was observed about 1 min later (Fig. 3, top). The mutant *rut¹*, *Ras1^{12a}/Ras1^{e2F}* or *raf^{HM7}* larvae showed no evoked PACAP38-like enhancement of K⁺ currents (Fig. 3).

As described earlier, activation of the cAMP pathway alone, achieved by application of cAMP analogues, failed to induce the PACAP38-like response (Fig. 1d). Activation of the Ras/Raf pathway alone was accomplished by using transgenic flies (*hsp70-raf^{gof}*), which carry an inducible transgene encoding a constitutively active (gain-of-function) Raf kinase (*Rag^{gof}*)²⁶. Perfusion of 8-bromo cAMP alone failed to induce the PACAP38-like response in wild-type larvae (Fig. 1d), heat-shocked wild-type larvae (not shown), uninduced (non-heat-shocked) *hsp70-raf^{gof}/+* (or *hsp70-raf^{gof}/TM3B*) transgenic larvae (Fig. 4a) or heat-shocked transgenic larvae carrying a *Shaker* K⁺ channel subunit (not shown)²⁷. Similarly, the PACAP38-like response was not observed in induced *hsp70-raf^{gof}/+* (or *hsp70-raf^{gof}/TM3B*) transgenic larvae (Fig. 4a, see 'control' in Fig. 4b).

In contrast to these negative results from activation of either signal transduction pathway alone, a combination of induction of *hsp70-raf^{gof}* and perfusion of 8-bromo cAMP in *hsp70-raf^{gof}/+* (or *hsp70-raf^{gof}/TM3B*) larvae produced a PACAP38-like response in a large fraction of muscle fibres tested (Fig. 4; see the legend for details). Enhancement of K⁺ currents occurred almost simultaneously with the inward current with 2 mM 8-bromo cAMP (Fig. 4b), but appeared 20 to 40 s later with 0.5 mM 8-bromo cAMP (not shown). Similar effects of 8-bromo

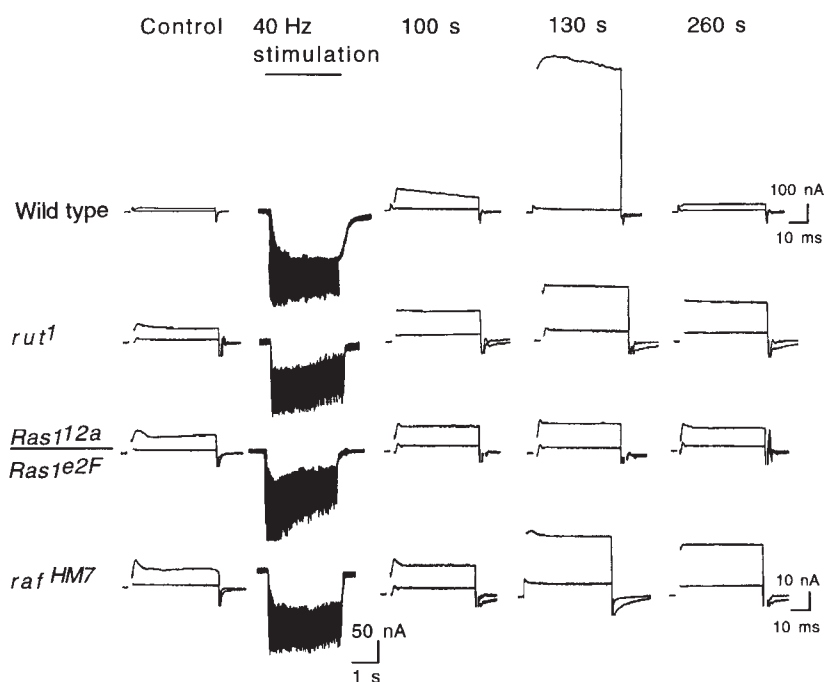
cAMP were also observed in heat-shock induced *rut¹*; *hsp70-raf^{gof}/+* and *raf^{HM7}*; *hsp70-raf^{gof}/+* double mutants (Fig. 4b).

Thus the PACAP38 response is diminished in *rut*, *Ras1* and *raf* mutants, and can be produced only by the concomitant expression of active Raf with application of cAMP analogues. Although other possibilities remain, the diminished PACAP38 response in these mutants probably results from defective Rut-AC, *Ras1* or *Raf*, while other aspects of PACAP38 activation remain largely intact.

This interpretation is supported by three observations: First, 1 μ M, but not 10 μ M, PACAP38 induces normal enhancement of K⁺ currents in the hypomorphic *Ras1^{12a}* mutants (see Fig. 2b, c). This suggests that all components of the PACAP38 signalling pathway can function normally in *Ras1^{12a}* mutants. The lack of response at 10 μ M PACAP38 (Fig. 2b), therefore, is likely to result from reduced *Ras1* activity in the mutant (C. Berg, personal communication). Because enhancement of K⁺ currents occurs 30 s after application of 10 μ M PACAP38, but 5 to 8 min later with 1 μ M PACAP38, it is possible that 10 μ M PACAP38 induces a rapid increase in the Rut-AC activity which is temporally uncoupled from slow activation of the *Ras1* pathway because of the reduced expression of *Ras1* in the *Ras1^{12a}* mutant. The preceding increase of cAMP might trigger an inhibitory mechanism that prevents enhancement of K⁺ currents. This cAMP-dependent inhibition is indicated by observations that application of 8-bromo-cAMP before PACAP38 blocked the PACAP38-induced enhancement of K⁺ currents (not shown), and that enhancement of K⁺ currents induced by active Raf and 8-bromo cAMP diminished even though 8-bromo cAMP persists (see Fig. 4b). When 1 μ M PACAP38 was applied, both *Ras1* and Rut-AC activities rise slowly and coincidentally, inducing the response. Second, coactivation of the Raf and cAMP pathways induces the PACAP38-like response in all genetic backgrounds tested (see Fig. 4). Third, this response could be induced as early as 30 min after heat shock in *raf^{HM7}*; *hsp70-raf^{gof}* (Fig. 4c) suggesting that no components downstream of Raf in the PACAP38 signalling pathway are disrupted developmentally in the *raf* mutant.

Hence coactivation of both the *Ras1*/Raf and Rut-AC pathways appears to mediate PACAP38-like neuropeptide trans-

FIG. 3 Evoked PACAP38-like enhancement of K⁺ currents is diminished in *rut*, *Ras1* and *raf* mutants. Stimulation (40 Hz) of wild-type motor axons for 3 s reliably induced a PACAP38-like enhancement of K⁺ currents in 1.5 mM Ca²⁺ (17 or 19 fibres). 'Control' currents were recorded before stimulation. During stimulation, muscle fibres were clamped to -80 mV. Evoked excitatory junctional current traces are shown under '40 Hz stimulation'. Then the same voltage paradigm for eliciting K⁺ currents as that used in the 'control' was repeated at different times (in seconds on top of current traces) after stimulation. Although a slow inward synaptic current was also evoked by high-frequency stimulation¹², a quantitative analysis of this synaptic current was prevented owing to difficulties in resolving it from the fast synaptic current gated by glutamate receptors³⁰. The reason for using higher concentrations of Ca²⁺ (1.5 mM instead of 1 mM) in these experiments was that it produced a more reliable enhancement of K⁺ currents with a shorter latency¹², presumably owing to the release of more peptides. Note the vertical scale for the wild-type current traces is different from that for the mutants; *n* = 17, 10, 8 and 10 for wild type, *rut¹*, *Ras1^{12a}/Ras1^{e2F}* and *raf^{HM7}*.



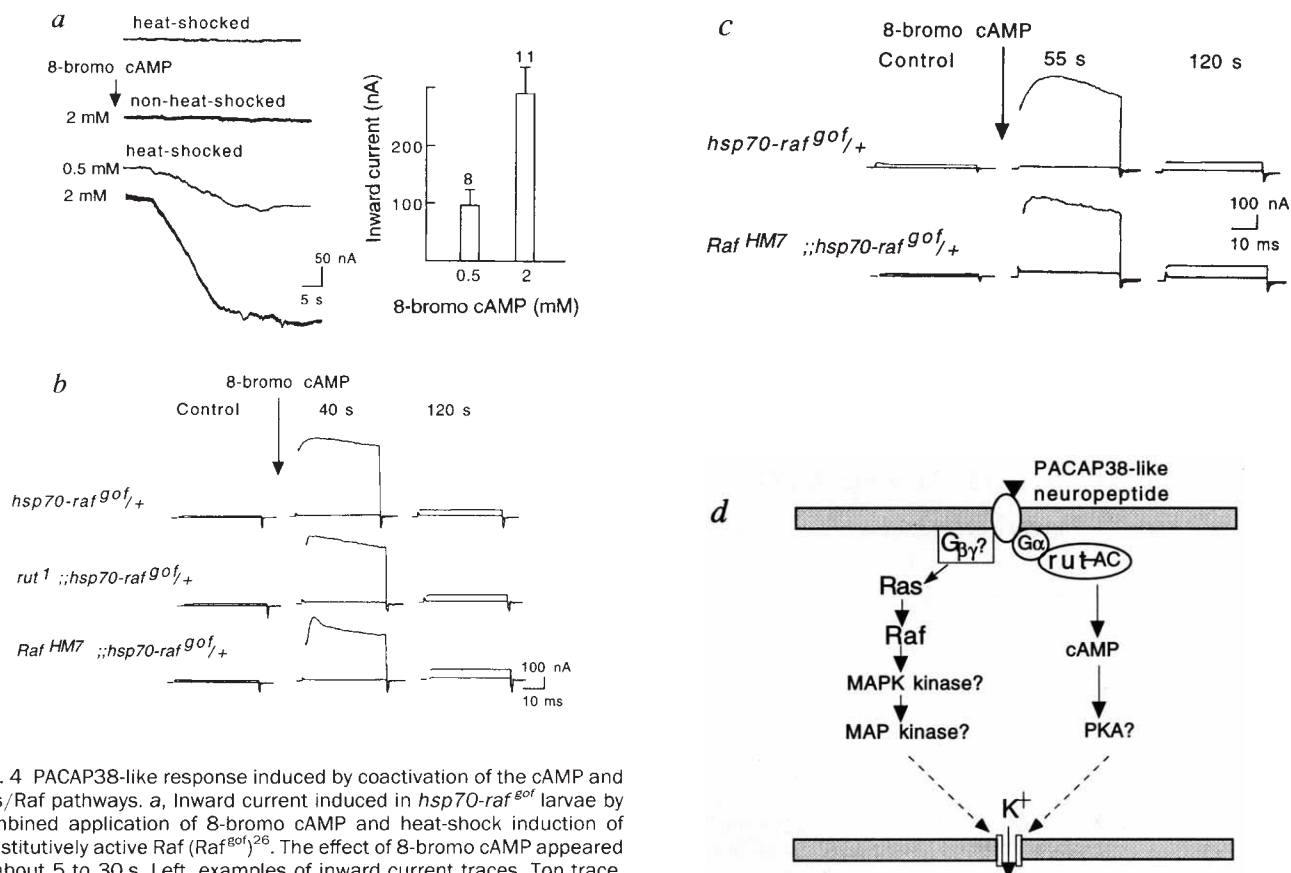


FIG. 4 PACAP38-like response induced by coactivation of the cAMP and Ras/Raf pathways. **a**, Inward current induced in *hsp70-raf^{gof}* larvae by combined application of 8-bromo cAMP and heat-shock induction of constitutively active Raf (Raf^{gof})²⁶. The effect of 8-bromo cAMP appeared in about 5 to 30 s. Left, examples of inward current traces. Top trace, activation of Raf kinase alone induced no inward current. Second trace, 2 mM 8-bromo cAMP failed to induce inward current in uninduced *hsp70-raf^{gof}* larvae. Bottom two traces, 8-bromo cAMP induced an inward current in heat-shocked *hsp70-raf^{gof}* larvae. Right, a histogram showing averaged peak inward current (mean $\pm$ s.e.m.); the number of muscle fibres that responded to 8-bromo cAMP are as indicated. The total number of muscle fibres examined was 15 for 0.5 mM and 19 for 2 mM 8-bromo cAMP. **b**, PACAP38-like enhancement of K⁺ currents induced by 8-bromo cAMP (2 mM) in heat-shocked *hsp70-raf^{gof}* larvae. Note that the PACAP38-like response diminished even though 8-bromo cAMP was presented persistently. About 60% of heat-shocked muscle fibres (35 out of 56 for *hsp70-raf^{gof/+}* and *hsp70-raf^{gof}/TM3*; 8 out of 17 for *rut1*; *hsp70-raf^{gof/+}*; 6 out of 10 for *raf^{HM7}*; *hsp70-raf^{gof}* larvae) responded to 8-bromo cAMP. It is possible that an inhibitory mechanism might be initiated in these non-responsive muscle fibres because they appeared also not to respond to PACAP38 (7 fibres examined). Larvae were subjected to a heat shock (37.5 °C) for 1 h or 1.5 h, and were dissected 1–2 h later. Heterozygous *hsp70-raf^{gof/+}* flies were bred from *hsp70-raf^{gof}/TM6B* and wild type (+/+). Hemizygous *rut1*; *hsp70-raf^{gof/+}* and *Raf^{HM7}*; *hsp70-raf^{gof/+}* flies were obtained from *rut1/rut1*; +/+ and +/Y; *hsp70-raf^{gof}/TM6B* or *raf^{HM7}/FM7*; +/+ and *FM7/Y*; *hsp70-raf^{gof}/TM6B* parents, respectively. **c**, Enhancement of K⁺ currents could be induced by active Raf and 8-bromo cAMP as early as 30 min after heat shock. Of 11 muscle fibres, 4 in *hsp70-raf^{gof}*

larvae, and 5 out of 14 in *raf^{HM7}*; *hsp70-raf^{gof}*, responded to 8-bromo cAMP (2 mM) in this time range. As controls, 14 fibres from wild-type and larvae carrying the *Shaker* transgene²⁷ failed to produce the enhancement of K⁺ currents tested 20–30 min after heat shock (not shown). **d**, A coincident signal transduction model for modulation of K⁺ currents induced by a PACAP38-like neuropeptide. Question marks indicate molecules whose function have not yet been tested. Dashed arrows indicate where the cAMP and Ras/Raf pathways may converge; whether such convergence is on channels directly or via an intermediate target remains to be determined. Consistent with the idea of coincident activation of the cAMP and Ras/Raf pathways are data from *Ras1^{12a}/Ras1^{12a}* mutants in Fig. 2b,c. When 10 μ M PACAP38 was applied, a rapid increase in the Rut-AC activity likely was temporally uncoupled from activation of the Ras1 pathway because of the reduced activity of Ras1 proteins in the mutant. This temporal mismatch of PACAP38-mediated activation of the two pathways then prevented the normal enhancement of K⁺ currents (Fig. 2b). In contrast, application of 1 μ M PACAP38 may activate both Ras and Rut-AC activities slowly but coincidentally, thereby allowing a normal enhancement of K⁺ currents (Fig. 2c).

mission. A model is proposed based on the data presented (Fig. 4d). Binding of a PACAP38-like peptide to its receptors leads to activation of Rut-AC by the G_α subunit and of Ras/Raf by the G_{βγ} complex; the pathways then converge to modulate K⁺ channel activity. This result will aid the study of the mechanisms and functions of G-protein-coupled neurotransmission mediated by peptides and other neuromodulators. □

Received 21 December 1994; accepted 6 April 1995.

- Nicoll, R. A. *Science* **241**, 545–551 (1988).
- Exton, J. H. A. *Rev. Physiol.* **56**, 349–369 (1994).
- Neer, E. J. *Cell* **80**, 249–257 (1995).
- Crespo, P., Xu, N., Simonds, W. F. & Gutkind, J. S. *Nature* **369**, 418–420 (1994).

- van Corven, E. J., Hordijk, P. L., Medema, R. H., Bos, J. L. & Moolenaar, W. H. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1257–1261 (1993).
- Egan, S. E. & Weinberg, R. A. *Nature* **365**, 781–783 (1993).
- Marshall, C. J. *Cell* **80**, 179–185 (1995).
- Herskowitz, I. *Cell* **80**, 187–197 (1995).
- Horvitz, H. R. & Sternberg, P. W. *Nature* **351**, 535–541 (1991).
- Zipursky, S. L. & Rubin, G. M. A. *Rev. Neurosci.* **17**, 373–397 (1994).
- Perrimon, N. *Cell* **74**, 219–222 (1993).
- Zhong, Y. & Pena, L. *Neuron* **14**, 527–536 (1995).
- Arimura, A. *Regul. Peptides* **37**, 287–303 (1992).
- Hille, B. *Ion Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1992).
- Spengler, D. et al. *Nature* **365**, 170–175 (1993).
- Livingstone, M. S., Szlober, P. P. & Quinn, W. G. *Cell* **37**, 205–215 (1984).
- Levin, L. R. et al. *Cell* **68**, 479–489 (1992).
- Han, P.-L., Levin, L. R., Reed, R. R. & Davis, R. L. *Neuron* **9**, 619–627 (1992).
- Neuman-Silberberg, F. S., Schejter, E., Hoffmann, F. M. & Shilo, B. Z. *Cell* **37**, 1027–1033 (1984).
- Simon, M. A. et al. *Cell* **67**, 701–716 (1991).

21. Yatani, A. et al. *Cell* **61**, 769–776 (1991).
22. Dickson, B., Sprenger, F., Morrison, D. & Hafen, E. *Nature* **360**, 600–603 (1992).
23. Rodriguez-Viciana, P. et al. *Nature* **370**, 527–532 (1994).
24. Martin, G. A. et al. *Science* **255**, 192–194 (1992).
25. Gauli, U., Mardon, G. & Rubin, G. M. *Cell* **68**, 1007–1019 (1992).
26. Brand, A. H. & Perrimon, N. *Genes Dev.* **8**, 629–639 (1994).
27. Zhao, M.-L., Sable, E. O., Iverson, L. E. & Wu, C.-F. *J. Neurosci.* **15**, 1406–1418 (1995).
28. Zhong, Y. & Wu, C.-F. *J. Neurogenet.* **9**, 15–27 (1993).
29. Stewart, B. A., Atwood, H. L., Renger, J. J., Wang, J. & Wu, C.-F. *J. comp. Physiol.* **A175**, 179–191 (1994).
30. Jan, L. Y. & Jan, Y. N. *J. Physiol. (Lond.)* **262**, 189–214 (1976).

ACKNOWLEDGEMENTS. I thank T. Tuily, R. Malinow, M.-M. Poo and N. Wright for critical comments on the manuscript, and G. Rubin, M. Simon, C. Berg, N. Perrimon and R. Davis for fly stocks. This work was supported, in part, by a grant from the Whitehall Foundation.

Preferential addition of newly synthesized membrane protein at axonal growth cones

Ann Marie Craig^{*†}, Russell J. Wyborski[‡]
& Gary Banker^{*}

^{*} Department of Neuroscience, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA

[‡] Biotechnology Department, Parke-Davis Research, Division of Warner-Lambert Co., Ann Arbor, Michigan 48106-1047, USA

THE addition of plasma membrane proteins to a growing axon could occur by preferential insertion at the tip (the growth cone), by uniform insertion along the axon, or by insertion at the cell body and bulk flow along the axon. To differentiate between these possibilities we used a defective herpesvirus vector to express an exogenous protein, the lymphocyte transmembrane protein CD8 α , in cultured rat hippocampal neurons. The newly synthesized protein first appeared on the axonal surface almost exclusively at the growth cone. Preferential addition at the growth cone was also observed in minor processes (immature dendrites), but not in mature dendrites. Over several hours, CD8 α reached a uniform distribution over the entire neuronal surface, presumably by diffusion within the membrane and possibly endocytic recycling. As well as providing materials for axonal growth, the selective addition of membrane vesicles at the growth cone may contribute to the polarized distribution of axonal surface molecules¹.

Contradictory views have arisen from previous studies that attempted to localize sites of membrane addition in growing axons^{2–8}. Instead of investigating bulk membrane movements, we determined the sites of addition of newly synthesized integral membrane proteins by inducing high-level expression of an exogenous plasma membrane protein in developing neurons and visualizing its surface distribution. Rat hippocampal neurons after 2 to 5 days in culture were infected with a defective herpes simplex virus vector expressing the lymphocyte transmembrane protein CD8 α . Herpes simplex viruses deliver the encapsidated DNA to the nucleus of infected cells, following which transcription and translation occur by normal cellular pathways⁹.

Protein expression from the herpesvirus-CD8 α vector was first detected in a few cells 8 hours after infection, and the number of expressing cells increased for several hours thereafter. Permeabilized cells exhibited strong CD8 α immunoreactivity in the cell body, at the distal end of the axon, and in small puncta resembling presumptive transport vesicles in the proximal and mid portions of the axon (Fig. 1a). When infected neurons were treated with brefeldin A, a drug which blocks transit of membrane proteins through the Golgi apparatus¹⁰, CD8 α staining

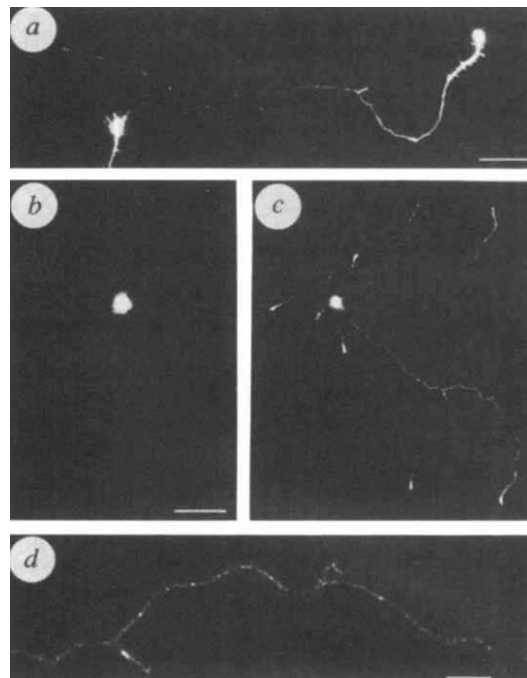


FIG. 1 Transit of CD8 α expressed from a defective herpesvirus vector through the Golgi complex and presumptive transport vesicles to the plasma membrane of cultured rat hippocampal neurons. *a*, Neurons were infected with a defective herpesvirus-CD8 α vector after 5 days in culture and fixed for immunostaining 12 h later. CD8 α -immunoreactivity was present at the distal end of the axon, in puncta resembling presumptive transport vesicles along the mid and proximal portions of the axon, and in the cell body (out of the field of view shown here). The absence of these puncta following surface labelling (see Fig. 2) confirmed that they are intracellular. *b–d*, Neurons were infected with the viral vector after 2 days in culture and co-incubated with brefeldin A, a drug that blocks transit of proteins through the Golgi complex¹⁰. *b*, In the presence of brefeldin A, the exogenous CD8 α accumulated within the cell body but did not enter the axon or minor processes. *c*, CD8 α -positive puncta appeared in the cell body, minor processes, and axon 90 min after removal of brefeldin (shown at higher magnification in *d*). CD8 α immunoreactivity was also present on portions of the plasma membrane (confirmed by surface labelling), particularly in the region of axonal and minor process growth cones. At even later times after removal of brefeldin A, surface labelling revealed the presence of CD8 α on the entire plasma membrane surface (not shown). Cultures that were not incubated with the viral vector exhibited no CD8 α immunoreactivity. Scale bars: *a*, 30 μ m; *b*, *c*, 50 μ m; *d*, 15 μ m.

METHODS. Neurons were obtained from 18-day embryonic rat hippocampi, plated on polylysine-coated glass coverslips, and maintained in serum-free medium above a glial feeder layer^{19,20}. The herpesvirus-CD8 α vector was prepared by construction of a CD8 α amplicon plasmid and packaging into virions, following methods developed and described elsewhere^{21–23}. The human CD8 α cDNA was isolated from plasmid pT8 (ref. 24) and cloned into the amplicon pF1'-P²⁵ under control of the HSV-1 α 4 promoter, producing plasmid pCD8314. E5 cells²⁶ (1×10^6) were transfected with 20 μ g pCD8314 using 40 μ l Transfectam (IBF) and were co-infected with d120 HSV-1 ICP4 deletion mutant helper virus²⁶ (1×10^7 plaque-forming units (PFU)). Viral progeny were collected and passaged 3 times on E5 cells, producing a virus stock with a helper virus titre of 4.2×10^5 PFU ml⁻¹ and a CD8 α -expressing viral vector titre of 9.7×10^5 virions ml⁻¹. Neuron-bearing coverslips were incubated with 3.5 μ l of this defective virus in 150 μ l medium for 1.5 h and returned to their home dishes. Brefeldin A (5 μ g ml⁻¹) was added for 10 h beginning 4 h after exposure of the cultures to the viral vector (before CD8 α expression was detectable). Neurons were fixed with 4% paraformaldehyde and permeabilized with 0.25% Triton X-100. CD8 α was detected by incubation with Mab DK25 (1:30; Dako) followed by biotinylated horse anti-mouse IgG (1:500; Vector) and then fluorescein isothiocyanate–streptavidin (1:500; Vector).

[†] Present address: Department of Cell and Structural Biology, University of Illinois, 506 Morrill Hall, 505 South Goodwin Avenue, Urbana, Illinois 61801, USA.

was restricted to the cell body (Fig. 1b). Shortly (50–90 min) after removal of the drug, CD8 α immunoreactivity also appeared in puncta along the neurites (Fig. 1c, d). At later times after removal of brefeldin A, CD8 α appeared on the entire plasma membrane surface of the neuron (not shown). Thus CD8 α expressed in neurons appears to be a suitable marker for following the transit of membrane proteins from their sites of synthesis in the endoplasmic reticulum through the Golgi complex to the plasma membrane.

Surface labelling at varying times after the initiation of expression was used to localize the sites where CD8 α was added to the plasma membrane (Fig. 2). Surface labelling was first detectable in a few cells 8 hours after exposure to virus. At this early time, the labelling was prominent only at the growth cones of axons and minor processes. Lighter labelling was present on the cell body, while little or no labelling was detectable along the length of the axon. Quantitation of the fluorescence signal indicated that the labelling at the midpoint of the axon was only 2–4% of that at the axonal growth cone. Between 8 and 12 hours after exposure to the viral vector, the number of labelled cells increased. In some cells at the 10 and 12 hour time points surface CD8 α was tightly concentrated at the growth cone, but in other cells strong labelling extended back from the growth cone towards the cell body. The simplest interpretation of this result is that the cells with label restricted to the growth cone correspond to cells that had just begun to express CD8 α on their surface, whereas the cells with label extending along the axon correspond to cells that had begun to express CD8 α at earlier times. By 20 hours after exposure to virus, the entire neuronal plasma membrane was strongly labelled, although the greatest intensity of labelling still tended to be at growth cones.

To quantify the changes in the pattern of CD8 α labelling over time, we measured the distance from the growth cone (where labelling was maximal) to the site along the axon where labelling

intensity had declined to 20% of that at the growth cone (Fig. 3). The average labelled distance increased with time, from 30 μ m at 8 h to 51 μ m at 10 h and 114 μ m at 12 h. During this period the number of labelled cells increased from 2.4 to 9 to 14 per coverslip. These values can be used to estimate the apparent rate of spread of CD8 α on the membrane surface. We calculate that about 17% (2.4 out of 14) of the cells labelled at 12 h must have begun to express CD8 α at 8 h. We assume that these correspond to the equivalent percentage of cells whose label extended the farthest back from the growth cone at 12 h. The average length of labelling in these cells was 257 μ m, which corresponds to a spread of labelling of 227 μ m in 4 h. Diffusion within the bilayer and addition at the tip with continued neurite elongation would both be expected to contribute to the spread of CD8 α . The expected distance of diffusion for a protein with a diffusion coefficient of 10^{-9} cm² s⁻¹, typical for a membrane protein, is 54 μ m in 4 h (ref. 1). The difference between this value and the observed rate of spread of CD8 α would imply a growth rate of 43 μ m h⁻¹, which is considerably faster than the previous estimate of 5 to 20 μ m h⁻¹ for cultured hippocampal neurons¹¹. It may be that an additional mechanism such as endocytic recycling contributes to the further spread of the protein in the membrane at later times.

Thus, by using a viral vector to induce a pulse of expression of CD8 α , we observed preferential addition of the newly synthesized membrane protein to axonal growth cones. This observation appears to contradict other results⁸ which suggest that most membrane components are added not at the growth cone but at the cell body and along the length of cultured *Xenopus* neurites. It is possible that CD8 α -containing vesicles deliver only a minor fraction of total membrane components and there is another route for delivery of bulk membrane. Alternatively, because other aspects of axon elongation, including growth rate, substrate attachment sites, and microtubule behaviour, differ

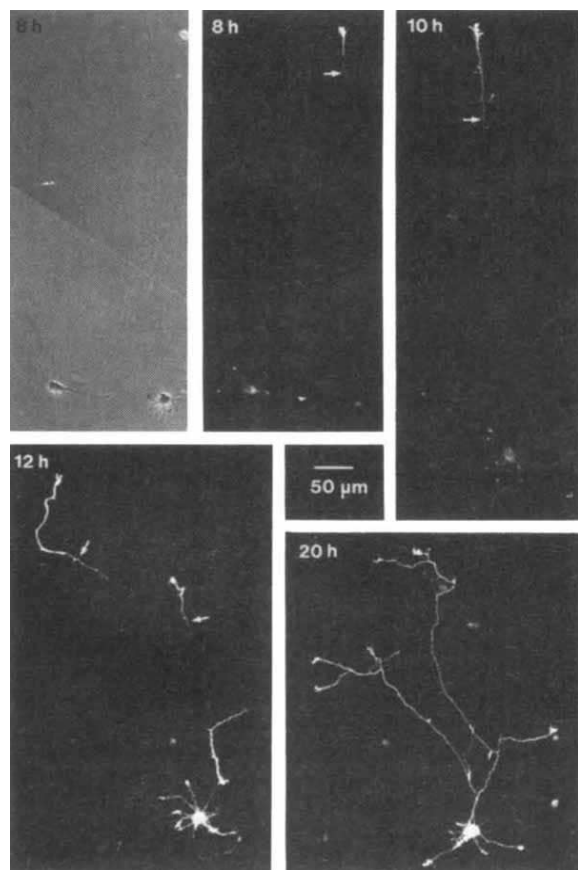


FIG. 2 Time course of the distribution of newly synthesized CD8 α on the plasma membrane of cultured rat hippocampal neurons. Neurons were incubated with the herpesvirus-CD8 α vector after 2 days in culture and immunostained for surface CD8 α at 8, 10, 12 and 20 h after exposure to the vector, as indicated. Surface immunoreactivity was first observed 8 h after exposure to the viral vector. At this early time, immunoreactivity was almost completely restricted to the growth cones of the axon and minor processes. The cell body was lightly labelled, and little or no immunoreactivity was present along the length of the axon. Note also the low background staining of the non-expressing neurons in this field. At 10 to 12 h after exposure to the vector, the neurons exhibited a range of immunostaining patterns for surface CD8 α . In some cells still exhibited strong surface labelling only at growth cones. In other cells, as in the examples shown here, strong surface labelling extended back from the growth cone along part of the distal axon, while the mid and proximal axon was still only very lightly labelled. In the 12 h cell shown here, the strong labelling that appears to be associated with the proximal portion of the axon represents staining of the distal portion of a branch of the axon that loops back towards the cell body. In other cells, the entire axon was significantly labelled, but the distal end was labelled most strongly. By 20 h after exposure to the vector, significant CD8 α labelling was present on the entire plasma membrane surface. The arrows indicate the point along the axon where fluorescence intensity was 20% of that at the growth cone (used for quantitation as described in Fig. 3).

METHODS. Cells were incubated with the herpesvirus-CD8 α vector after 2 days in culture, as described in Fig. 1. In some experiments, neurons were immunolabelled for surface CD8 α following fixation without permeabilization. In other experiments, live cells were incubated with the mAb DK25 (1:25 in medium) for 20 min, rinsed and then fixed and incubated in secondary antibody and fluorescein isothiocyanate-streptavidin. Similar patterns of CD8 α immunoreactivity were obtained with both methods.

between mammalian and *Xenopus* neurites^{2,8,12,13}, the sites of addition of membrane components may also differ. Elongation of mammalian axons may be more dependent on assembly of cytoskeletal elements and insertion of new membrane constituents at the tip.

The preferential addition of CD8 α at the growth cone resulted in a distal to proximal gradient of CD8 α protein along the axonal surface that was maintained for several hours after the initiation of expression. Although the rate at which this gradient dispersed was faster than had been predicted¹, our observation confirms the prediction¹ that preferential insertion of a membrane protein at the tip of an axon could generate its polarized distribution.

The growth cones of minor processes, the precursors of dendrites, were also frequently labelled preferentially for CD8 α . In more mature neurons that had been in culture for 14 days we observed preferential addition of CD8 α to axonal growth cones, but never to dendritic growth cones. In mature dendrites, as soon as we observed CD8 α on the surface, it was present along the entire dendrite. Although we cannot rule out the possibility that vesicles may also be selectively added to dendritic growth cones, the presence of microtubules of both polarity orientations in dendrites but not in axons or minor processes^{14,15} could result in a difference between axons and dendrites in the sites where vesicles are added.

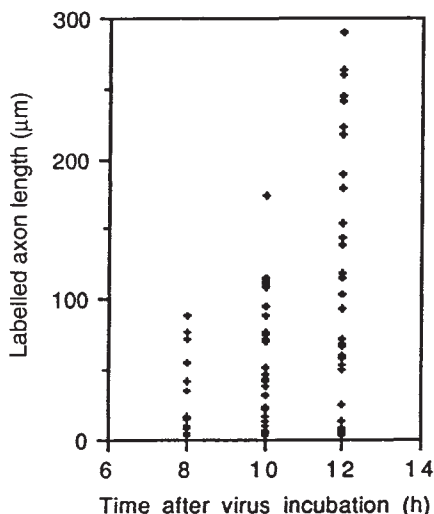


FIG. 3 Spread of CD8 α immunoreactivity along the axon at varying times after exposure of neurons to a defective herpesvirus-CD8 α vector. Rat hippocampal neurons were incubated with the herpesvirus-CD8 α vector after 2 days in culture and immunostained for surface CD8 α at 8, 10 or 12 h after exposure to the viral vector. At all three time points, immunofluorescence intensity along the axon was maximal at the growth cone and decreased proximally (Fig. 2). The distance from the maximal immunofluorescence at the growth cone to the point on the axon where immunofluorescence intensity dropped to 20% of the maximum was taken as a measure of the length of labelling. This corresponds approximately to the point where the signal drops off most sharply by visual inspection (Fig. 2). By this criterion, the average labelled lengths were 30 μ m at 8 h ($n=15$), 51 μ m at 10 h ($n=32$), and 114 μ m at 12 h ($n=31$). For 3 of the axons at the 12 h time point, this length corresponded to the total length of the axon. Average axon length was 270 μ m.

METHODS. Induction of CD8 α expression and immunolabelling were performed as described in Figs 1 and 2. Coverslips were scanned on a Zeiss Axiovert microscope and images of all immunoreactive neurons were recorded with a Hamamatsu SIT camera and a Panasonic optical disc recorder. Measurements were performed using Image 1 software (Universal Imaging). Each axon branch was analysed separately, and only those branches with growth cones visible by phase-contrast microscopy were included for analysis. These measurements were obtained from two separate experiments; similar patterns of labelling were observed in two other independent experiments but these were not quantified.

There are likely to be multiple populations of transport vesicles that deliver different components from the Golgi complex to the neuronal plasma membrane^{16–18}. We do not know what percentage of these contain CD8 α , nor do we know if different populations of vesicles are added to the plasma membrane at different sites. Expression of other proteins from viral vectors will probably be useful in identifying populations of transport vesicles and following their trafficking. □

Received 22 February; accepted 21 March 1995.

1. Futerman, A. H., Khanin, R. & Segel, L. A. *Nature* **362**, 119 (1993).
2. Bray, D. *Proc. natn. Acad. Sci. U.S.A.* **65**, 905–910 (1970).
3. Koda, L. Y. & Partlow, L. M. *J. Neurobiol.* **7**, 157–172 (1976).
4. Feldman, E. L., Axelrod, D., Schwartz, M., Heacock, A. M. & Agranoff, B. W. *J. Neurobiol.* **12**, 591–598 (1981).
5. Pfenninger, K. H. & Maylie-Pfenninger, M.-F. *J. Cell Biol.* **89**, 547–559 (1981).
6. Griffin, J. W., Price, D. L., Drachman, D. B. & Morris, J. J. *Cell Biol.* **88**, 205–214 (1981).
7. Small, R. K., Blank, M., Ghez, R. & Pfenninger, K. H. *J. Cell Biol.* **98**, 1434–1443 (1984).
8. Popov, S., Brown, A. & Poo, M.-M. *Science* **259**, 244–246 (1993).
9. Roizman, B. & Sears, A. E. in *Virology* Vol. 2 (eds Fields, B. N. et al 1795–1841 (Raven, New York, 1990).
10. Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. *J. Cell Biol.* **116**, 1071–1080 (1992).
11. Dotti, C. G., Sullivan, C. A. & Banker, G. A. *J. Neurosci.* **8**, 1454–1468 (1988).
12. Reinsch, S. S., Mitchison, T. J. & Kirschner, M. J. *Cell Biol.* **115**, 365–379 (1991).
13. Okabe, S. & Hirokawa, N. *J. Cell Biol.* **117**, 105–120 (1992).
14. Baas, P. W., Deitch, J. S. & Black, M. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8335–8339 (1988).
15. Baas, P. W., Black, M. M. & Banker, G. A. *J. Cell Biol.* **109**, 3085–3094 (1989).
16. Dotti, C. G. & Simons, K. *Cell* **62**, 62–72 (1990).
17. Jung, L. J. & Scheller, R. H. *Science* **251**, 1330–1335 (1991).
18. Mundigl, O. et al. *J. Cell Biol.* **122**, 1207–1221 (1993).
19. Banker, G. A. & Cowan, M. W. *Brain Res.* **126**, 397–425 (1977).
20. Goslin, K. & Banker, G. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 207–226 (MIT Press, Cambridge, MA, 1991).
21. Spaete, R. R. & Frenkel, N. *Cell* **30**, 295–304 (1982).
22. Geller, A. I. & Breakefield, X. O. *Science* **241**, 1667–1669 (1988).
23. Kaplitt, M. G. et al. *Molec. cell. Neurosci.* **2**, 320–330 (1991).
24. Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L. & Axel, R. *Cell* **40**, 237–246 (1985).
25. Kwong, A. D. & Frenkel, N. *Virology* **142**, 421–425 (1985).
26. DeLuca, N. A., McCarthy, A. M. & Schaffer, P. A. *J. Virol.* **56**, 558–570 (1985).

ACKNOWLEDGEMENTS. We thank D. I. Gottlieb for his contribution to the initial stages of this work; E. M. Johnson, P. Olivo and colleagues for advice on the virus methodology; N. Frenkel and D. Littman for the plasmids pF1-P and pT8; H. Asmussen and H. Edelin for technical assistance; and A. H. Futerman for comments on the manuscript. This work was supported by grants from the NIH.

Ca²⁺-dependent and -independent activities of neural and non-neural synaptotagmins

Cai Li*, Beate Ullrich*, Jimmy Z. Zhang†, Richard G. W. Anderson†, Nils Brose* & Thomas C. Südhof*‡§

Departments of * Molecular Genetics and † Cell Biology and Neuroscience, and ‡ Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA

SYNAPTOTAGMINS (Syts) are brain-specific Ca²⁺/phospholipid-binding proteins^{1–5}. In hippocampal synapses, Syt I is essential for fast Ca²⁺-dependent synaptic vesicle exocytosis but not for Ca²⁺-independent exocytosis³. In vertebrates and invertebrates^{6–9}, Syt may therefore participate in Ca²⁺-dependent synaptic membrane fusion, either by serving as the Ca²⁺ sensor in the last step of fast Ca²⁺-triggered neurotransmitter release, or by collaborating with an additional Ca²⁺ sensor. While Syt I binds Ca²⁺ (refs 10, 11), its phospholipid binding is triggered at lower calcium concentrations (EC₅₀ = 3–6 μ M) than those required for exocytosis¹². Furthermore, Syts bind clathrin-AP2 with high affinity, indicating that

§ To whom correspondence should be addressed.

TABLE 1 Differential properties of synaptotagmins

	Tissue distribution	Ca ²⁺ -dependent binding of phospholipids		Ca ²⁺ -dependent binding of syntaxin		Ca ²⁺ -independent binding of AP2
		C ₂ -A	C ₂ -A/B	C ₂ -A	C ₂ -A/B	
Syt I	Rostral brain	3–6 μ M	3–6 μ M	200–400 μ M	200–400 μ M	0.1–1.0 nM
Syt II	Caudal brain	3–6 μ M	ND	300–500 μ M	ND	ND
Syt III	Brain	3–6 μ M	ND	< 10 μ M	ND	0.1–1.0 nM
Syt IV	Brain (ubiquitous)	No binding	ND	No binding	ND	0.1–1.0 nM
Syt V	Brain	3–6 μ M	ND	300–500 μ M	ND	0.1–1.0 nM
Syt VI	Ubiquitous	No binding	ND	No binding	Ca ²⁺ -independent	0.1–1.0 nM
Syt VII	Ubiquitous	3–6 μ M	ND	< 10 μ M	< 10 μ M	0.1–1.0 nM
Syt VIII	Ubiquitous	No binding	ND	No binding	ND	0.1–1.0 nM

ND, not determined; C₂-A, first C₂ domain; C₂-A/B, both C₂ domains of a synaptotagmin. Tissue distribution was assessed by PCR, RNA blots and immunoblots (Fig. 1 and data not shown). PCRs suggest significant but low-level ubiquitous expression of synaptotagmin IV (data not shown). For Syt IV, VI and VII, mRNA levels are much higher in brain than in other tissues. Phospholipid and syntaxin binding was measured in the presence of 0.1 and 0.15 M NaCl, respectively, as a function of Ca²⁺, to the first C₂ domains of the indicated Syts fused to GST; values indicate half-maximal effective concentrations (EC₅₀) (Figs 2 and 3). Ca²⁺-dependent phospholipid and syntaxin binding to the double C₂ domain fragment of Syt I was measured using brain proteins or purified Syt I subjected to mild trypsin treatment which cleaves Syt I selectively at a single hypersensitive site N-terminal to the first C₂ domain (B. Davletov and T.C.S., unpublished results)²¹. C₂-A/B binding to syntaxin for Syt VI and VII was measured $\pm$ 1 mM Ca²⁺ to recombinant Syts (Fig. 3); in contrast to the single C₂ domain, some binding is observed to double C₂ domains in the absence of Ca²⁺. AP2 binding was measured as described^{4,5} to the double C₂ domains of the indicated synaptotagmins. In the case of Syt I (ref. 4) and Syt VIII (this study) the binding site was localized to the second C₂ domain and is likely to be in that domain for all Syts. Values shown are approximate K_d values for AP2 and they varied between experiments. The data on the distribution, the AP2 and phospholipid binding for Syts I, II, III and IV are from refs 4, 5, 13–16.

they may play a general role in endocytosis^{4,5} rather than being confined to a specialized function in regulated exocytosis³. Here we resolve this apparent contradiction by describing four Syts, three of which (Syt VI, VII and VIII) are widely expressed in non-neural tissues. All Syts tested share a common domain structure, with a cytoplasmic region composed of two C₂ domains that interacts with clathrin-AP2 (K_d=0.1–1.0 nM) and with neural and non-neural syntaxins. The first C₂ domains of Syt I, II, III, V and VII, but not of IV, VI or VIII, bind phospholipids with a similar Ca²⁺-concentration dependence (EC₅₀=3–6 μ M). The same C₂ domains also bind syntaxin as a function of Ca²⁺ but the Ca²⁺-concentration dependence of Syt I, II and V (>200 μ M) differs from that of Syt III and VII (<10 μ M). Syts therefore appear to be ubiquitous proteins with a role in exocytosis mediated by syntaxin binding. The Ca²⁺ levels needed to trigger syntaxin binding by the different Syts suggest that they play distinct roles in membrane fusion; the level required by Syt I approximates those required for synaptic exocytosis.

Four new Syts were cloned. All four new Syts are rare, as judged by the abundance of complementary DNA clones, and only partial sequences were isolated for two of the four messenger RNAs. Alignment of the amino-acid sequences of the new Syts with those of Syt I–IV^{13–16} shows that they share the domain structure common to all Syts, including an intravesicular N terminus without a signal sequence, a single transmembrane region, a linker sequence and two C₂ domains (Fig. 1a). RNA blots, polymerase chain reaction (PCR) and immunoblots (Fig. 1b–d, and data not shown) confirmed that Syt I, II and III are almost exclusively expressed in neural tissues¹⁴, as is the new Syt V. The other three new Syts, however, are present in most tissues examined. Although levels of Syt VI and VII are highest in brain, Syt VII is also abundant in heart and lung (Fig. 1b and Table 1, and data not shown). Immunoblots with affinity-purified antibodies confirmed that Syt VI and VII protein is present (Fig. 1d).

The first C₂ domains of Syt I, II and III bind phospholipids as a function of Ca²⁺, and the second C₂ domains of Syt I, III and IV bind AP2 independently of Ca²⁺ (Syt II has not been tested)⁵. Syt I also interacts with the cytoplasmic domains of neurexins^{17,18} and with syntaxin I, a plasma membrane protein required for synaptic vesicle exocytosis^{19,20}. Although Ca²⁺ binding induces a conformational change in Syt I (ref. 21), its protein–protein interactions are not Ca²⁺-dependent and thus cannot explain its role in fast Ca²⁺-triggered vesicle exocytosis³.

The first C₂ domains of Syt V and VII bound phospholipids as a function of Ca²⁺ (Fig. 2), unlike those of Syt VI and VIII. Their Ca²⁺-dependence was similar to that of Syt I, II and III. All Syts bound AP2 with a high affinity in the absence of Ca²⁺ (Table 1 and data not shown); immunoblotting confirmed that AP2 was bound selectively but not AP1 or β -COP. Thus, neural and non-neural Syt isoforms differ in their Ca²⁺-dependent phospholipid binding but bind clathrin-AP2 with a uniformly high affinity (summarized in Table 1).

During synaptic exocytosis, syntaxin I forms a complex with SNAP-25 (synaptosome-associated protein, M_r 25K) and synaptobrevin/VAMP (for vesicle-associated membrane protein) (the 'core complex') which is required for fusion^{22–24}. The core complex serves as a receptor for NSF and SNAPS, ubiquitous membrane fusion proteins. Syt I competes with SNAPS for binding to the core complex²², indicating that Syt I binding to syntaxin I may be a step in the fusion reaction. Syntaxin and synaptobrevin homologues are widely distributed in non-neural tissues, where they are thought to act as SNAP receptors (SNAREs) in various membrane fusion reactions^{25,26}, suggesting that the non-neural Syts may serve a ubiquitous function in membrane fusion by interaction with non-neural syntaxins. We therefore examined the interactions of ubiquitous and synaptic Syts with neural and non-neural syntaxins (Fig. 3a). These studies were performed in the presence of Mg²⁺ with or without Ca²⁺ by binding recombinant Syt I, VI and VII to recombinant syntaxins I to IV, recombinant Syt I to VIII to endogenous brain syntaxin I, and native brain Syt I to recombinant syntaxins I to IV.

Purified syntaxins I to IV fused to maltose-binding protein (MBP-syntaxins) were incubated with immobilized recombinant glutathione S-transferase (GST)–Syt I, VI or VII containing either the first C₂ domain or both C₂ domains. All syntaxins bound to the two C₂ domains of the three Syts, indicating that like SNAP-25 and Munc18, Syts bind to all four syntaxins (Fig. 3a and data not shown). Ubiquitous Syts may therefore interact with non-neural syntaxins in non-synaptic membrane fusion just as Syt I is thought to interact with syntaxin I during synaptic exocytosis. As high concentrations of recombinant Syts and syntaxins were used, the binding observed does not reflect the relative affinities between different syntaxins and Syts. At limiting concentrations, Syt I interacts more tightly with syntaxin I than with syntaxins II, III or IV (data not shown), suggesting that at physiological concentra-

Syt I	MVSASHPEALAAAPVTTVATLVPHNATEPASPGEGKEDAFSLKQKQFMNELHKIPL	PPWALIIAIAIVAV	68
Syt II	MRNIFKRNQEPVAPATTTATMPLAPAAPADNSTESTGTGESQEDMPAKLKDKFFNEINKIPL	PPWALIIAMAVVAG	76
Syt III	MSGDYEDDLCCRALLILVSDLCARIRDADTNDRCQEFNELRIRGYPRGPDADIS	VSLLSVIVTFCCI	66
Syt IV	MAPITTSRVEFDEIPTVVGIFSAFGLVFT		29
Syt VI	MSGVWGAGGPRCQAALAVLASLCRARPPPLGLDVTETCSQFELQPPPEQSPSAADSGTS	VSLLAIVVIVCGV	70
Syt VII	MYRDPRAASPGAPTRDVLV	VSATITVSLSVTI	32
Syt VIII	...PRWTL	FIAILAAGVLLVS	18
Syt I	LVTTCFCFCCKKCLFKKKNKKKGKKGKGNAINMKDVKDLGKTMKDQALKDDDAETGLTDGEEK		133
Syt II	LLLTCCPCICCKKCCCKKKNKKKGKKGKGNAINMKDVKDLGKTMKDQALKDDDAETGLTDGEEK		134
Syt III	VLGVSLFVSWKLCWVWPWRDGGSAVGGGLRKLDAAGVGLAGLVGGGGHHLGASLGGHPLLGPPHHHAHPAHHP		142
Syt IV	VSFAWICCCQRRRSASKNKTPPYKPVHVLKGVDIYPENLSSKKKFGGDDKSEAKRKAALPNLSLHLDLEKRDLDNGNF		105
Syt VI	AVAVVFFFLFQKLCWMPWRNKKEASSPSSANPASEILOSPSSSRGNMADKLDKPSALGFLAEAAVKSISTSPDPAEVQ		146
Syt VII	VLCGLCHWCQRKRLGKRYKNSLETVGTPTDSGRGRGEKKAIKLPAAGKXAVNTAPVPGQTPHDES		108
Syt VIII	CLLCVICCYCHRRHRHRKQPKDKETVGLGSARNSTTTHLVQPDVDCLE		65
Syt III	FAELLEFGGLGGSEPPPSYLDMDSYPEAAVASVVAAGVKPSQTSPELPSEGGTGSGLLLLPSSGGGLPSA		213
Syt VI	PKTNPKAGSSDLENTPKLPETETKEAVSFESLKSSTSL		145
Syt VII	MSVKEHIMRHTKLQROTTEPASSTRHTSFKRHLRPMHVSSVDYGNELPPAAAEQPTSIGRIKPELYKQKSVGDDE		222
Syt VIII	NSLTSEHMLSLSPGSEDEA		127
Syt III	QSHQVTS LAPTTTRYPALPRPLTQQTTLTQADPSSEERPPALPLFLPLPGGEEKAKLIGQIKPELYQGTGFGGRTTG		289
Syt I	BEKPEEELKGLKLYSLDYDFQNNQQLVGVIIQAAELPALDM-GSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		208
Syt II	BEKPEEENLGLKLYSLDYDFQNNQQLVGVIIQAAELPALDM-GSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		209
Syt III	GSAGAGAPCGRISFALRYLYGSDQQLVVRILQALDLPAKDS-NGSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		364
Syt IV	TSEKQELGLTLFLSLLEVNFEEKAFVVRNIKEAAGLPAKDEQSMSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		221
Syt V	GRSSSKACGKGLFIFILKVDYDFQNNQQLVGVIIQAAELPALDM-GSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		75
Syt VI	AKSEAAKSCGKINFSRLRYLYGSDQQLVVRILQALDLPAKDS-NGSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		297
Syt VII	HEGCGENLGRIGRIQSVNFGSSTLTIVKVMKAQNLPAKDF-SGSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		202
Syt VIII	PCSGGQDQWGLLLSLLEVNFEEKAFVVRNIKEAAGLPAKDEQSMSTSDPYVVKVFLLPDKKKKFETKVRHRTLNPFVNE		137
Syt I	QTFK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		281
Syt II	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		282
Syt III	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		438
Syt IV	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		296
Syt V	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		149
Syt VI	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		371
Syt VII	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		275
Syt VIII	TFTEK-VPYSELGGKTLVMAVYDFDRFSKHDIIGEFKVV-FMNTW-DFGHVTEEWDRDLQSAEKKEEQKRLGDCIFSLR		210
Syt I	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		357
Syt II	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		358
Syt III	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		514
Syt IV	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		372
Syt V	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		225
Syt VI	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		447
Syt VII	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		351
Syt VIII	VVPTAGKLTIVVILEAKNKLKMDVGGGLSDPYVVKVHLMQNGKRLKKKKKTITIKNTLNPFYNEASFSEFVEFBIQKQVQV		282
Syt I	VVTVLVDYDKIGKNDADIKVFFVGYNSTGAELR-HWS DMLANPRRPIAQWHTLQVEEEDVDAMLA VKK		421
Syt II	VVTVLVDYDKIGKNDADIKVFFVGYNSTGAELR-HWS DMLANPRRPIAQWHTLQVEEEDVDAMLA VKK		422
Syt III	SFAVVDYDYGKNEAIGVCRVGPAAADPHGREHWAEMLAANPRRPIAQWHTLQVEEEDVDAMLA VKK		588
Syt IV	BFLVLDSEGRSRNEVIGRLVLGATAEG-SGGGHWSKEICDFPRRQIAKWHMLCDG		425
Syt V	SFAVVDYDYGKNEAIGVCRVGPAAADPHGREHWAEMLAANPRRPIAQWHTLQVEEEDVDAMLA VKK		279
Syt VI	LHSMVDYDYGKNEAIGVCRVGPAAADPHGREHWAEMLAANPRRPIAQWHTLQVEEEDVDAMLA VKK		511
Syt VII	ITVMDYDYGKNEAIGVCRVGPAAADPHGREHWAEMLAANPRRPIAQWHTLQVEEEDVDAMLA VKK		403
Syt VIII	VLAVWARGLQLRTEPVGKVLGSRASG-QPLQHWADMLAHARRPIAQWHTLQVEEEDVDAMLA VKK		355

FIG. 1 a, Primary structures of synaptotagmins. Sequences are shown in single-letter amino acid code aligned for maximal homology. Residues present in at least 50% of the sequences are shown on a black background. The putative transmembrane region is boxed. All sequences are from rat, except for Syt VIII which is from mouse. Syt V and Syt VIII sequences are incomplete at the N terminus. DNA sequences have been deposited in GenBank (accession numbers U20105 to U20110). b, Tissue distribution of Syt VII (arrow) analysed by RNA blotting. A blot containing poly(A)⁺-enriched RNA from the indicated tissues was hybridized with a Syt VII cDNA probe and

exposed for 16 h (left) or 80 h (right, to visualize low-level expression in skeletal muscle, kidney and testis). c, Expression of Syt mRNAs in rat tissues assayed by RT-PCR. Expression of cellubrevin²⁴ (Ceb, positive control) and of Syts was examined by performing PCRs with specific oligonucleotides on cDNA synthesized from total RNA of the indicated tissues. Products were analysed by Southern blotting using internal probes. The rightmost lane in all samples contains a control in which reverse transcriptase was omitted. All PCRs cover regions of the mRNA that are encoded by multiple exons separated by introns. Upper bands observed in some PCRs are mostly

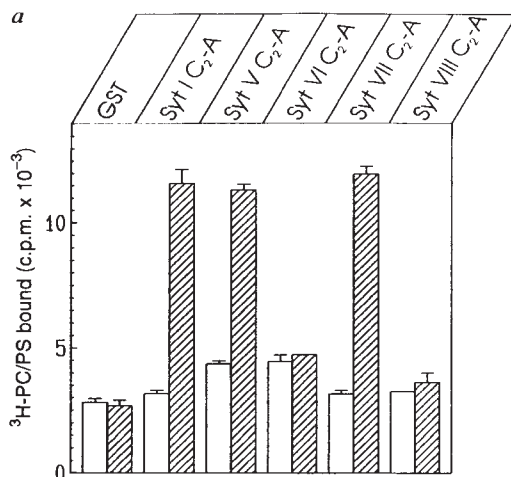
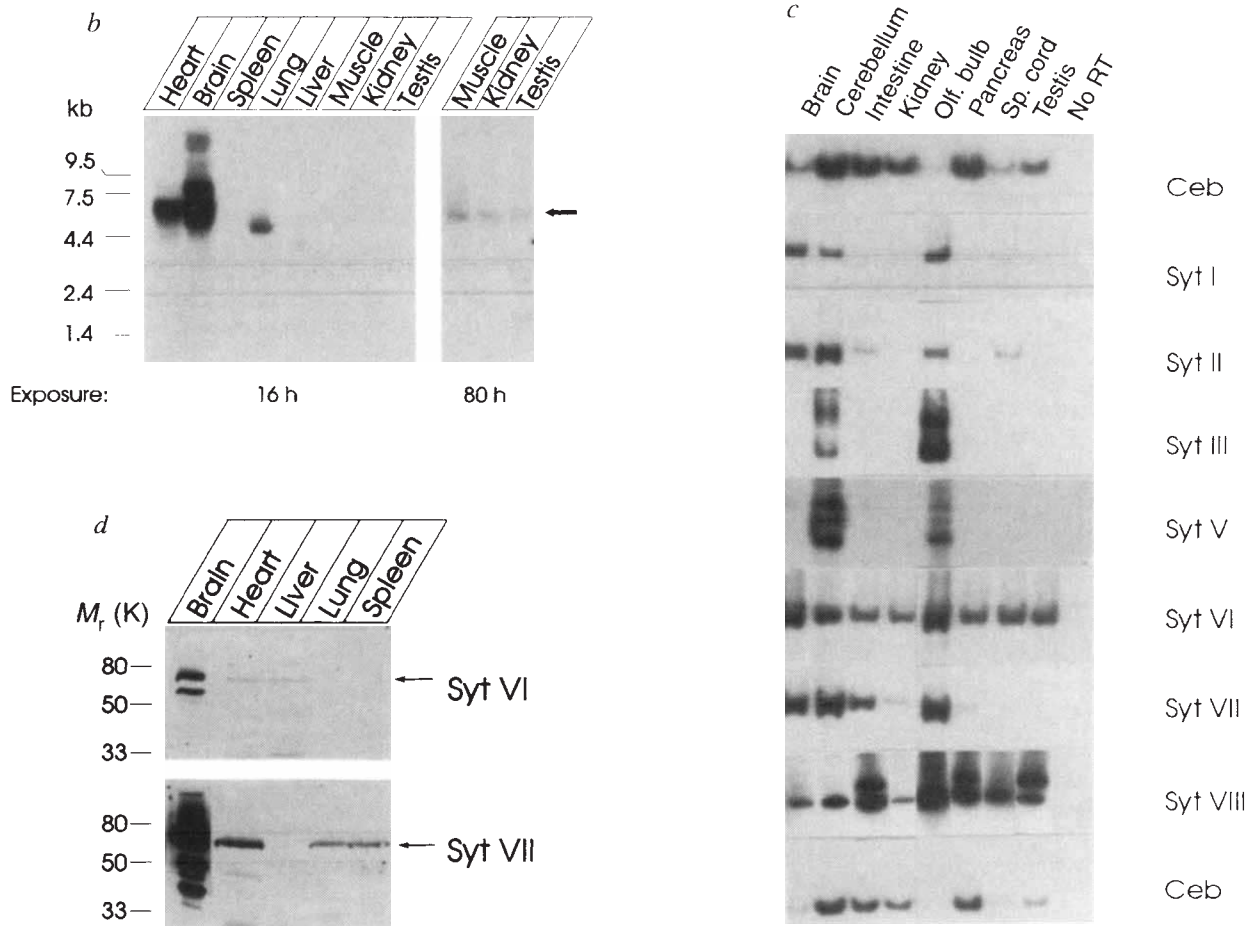


FIG. 2 a, Ca²⁺-dependent phospholipid binding to the first C₂ domain of Syts. GST and GST-Syt fusion proteins were attached to glutathione-agarose beads and incubated with ³H-labelled liposomes in the absence (open bars) or presence (hatched bars) of 0.1 mM Ca²⁺. Beads were washed and bound radioactivity was counted. b, Ca²⁺-affinity for phospholipid binding by Syt I, V and VII. Phospholipid binding to the first C₂ domains of the indicated proteins was measured as a function of the free Ca²⁺ concentration.

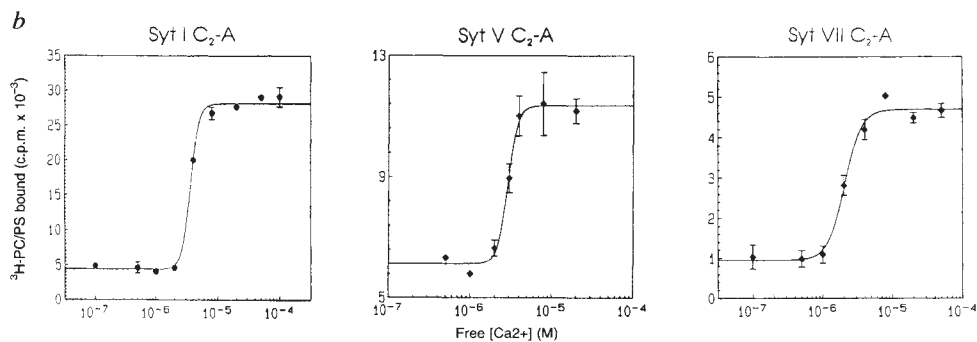
METHODS. GST and GST-fusion proteins were produced from recombinant pGEX-KG plasmids²⁹ constructed by amplifying the part of the coding region to be expressed by PCR, or by use of existing restriction sites. GST-fusion proteins used in Figs 2 and 3 and Table 1 contain the following amino acids (Fig. 1a): Syt I C₂-A, 140–267; Syt I C₂-B, 265–421; Syt II C₂-A, 141–273; Syt II C₂-B, 272–422; Syt III C₂-A, 293–425; Syt IV C₂-A, 152–282; Syt V C₂-A, 6–136; Syt VI C₂-A, 228–358; Syt VI C₂-A/B, 228–511; Syt VII C₂-A, 134–262; Syt VII C₂-A/B, 134–403; Syt VIII C₂-A, 57–198; Syt VIII C₂-B, 198–355. Phospholipid binding assays were performed with Ca²⁺/EGTA buffers and phosphatidylcholine/phosphatidylserine liposomes (2.5:1)⁵. Error bars indicate standard deviations from triplicate determinations.



from contaminating genomic DNA in the sample. *d*, Immunoblot of proteins from rat tissues with affinity-purified Syt VI and VII antibodies. Membrane proteins enriched by Triton X-114 extraction from the indicated tissues were separated by SDS-PAGE and immunoblotted with affinity-purified antibodies.

METHODS. Genomic clones encoding parts of Syt VI and VIII genes were isolated by low-stringency screening of murine genomic libraries; exonic probes were then used to isolate cDNAs encoding Syt III, IV, V, VI and VII from rat liver, thymus, intestine and brain cDNA libraries^{3,5,14,28}. The Syt VIII genomic clone contained all of the coding

region shown; exon-intron junctions were confirmed by sequencing PCR products obtained with cDNA from mouse and rat tissues. Mapping, sequencing and expression of clones and PCR products were performed as described^{3,5,14,27,28}. RNA blots were obtained from Clontech and hybridized at high stringency¹⁴. Antibodies against Syt VI and Syt VII were affinity-purified on immobilized Syt VI and VII GST-fusion proteins. Specificity of antibodies was tested using COS cells transfected with full-length Syt I, II, III, IV, VI and VII constructs.



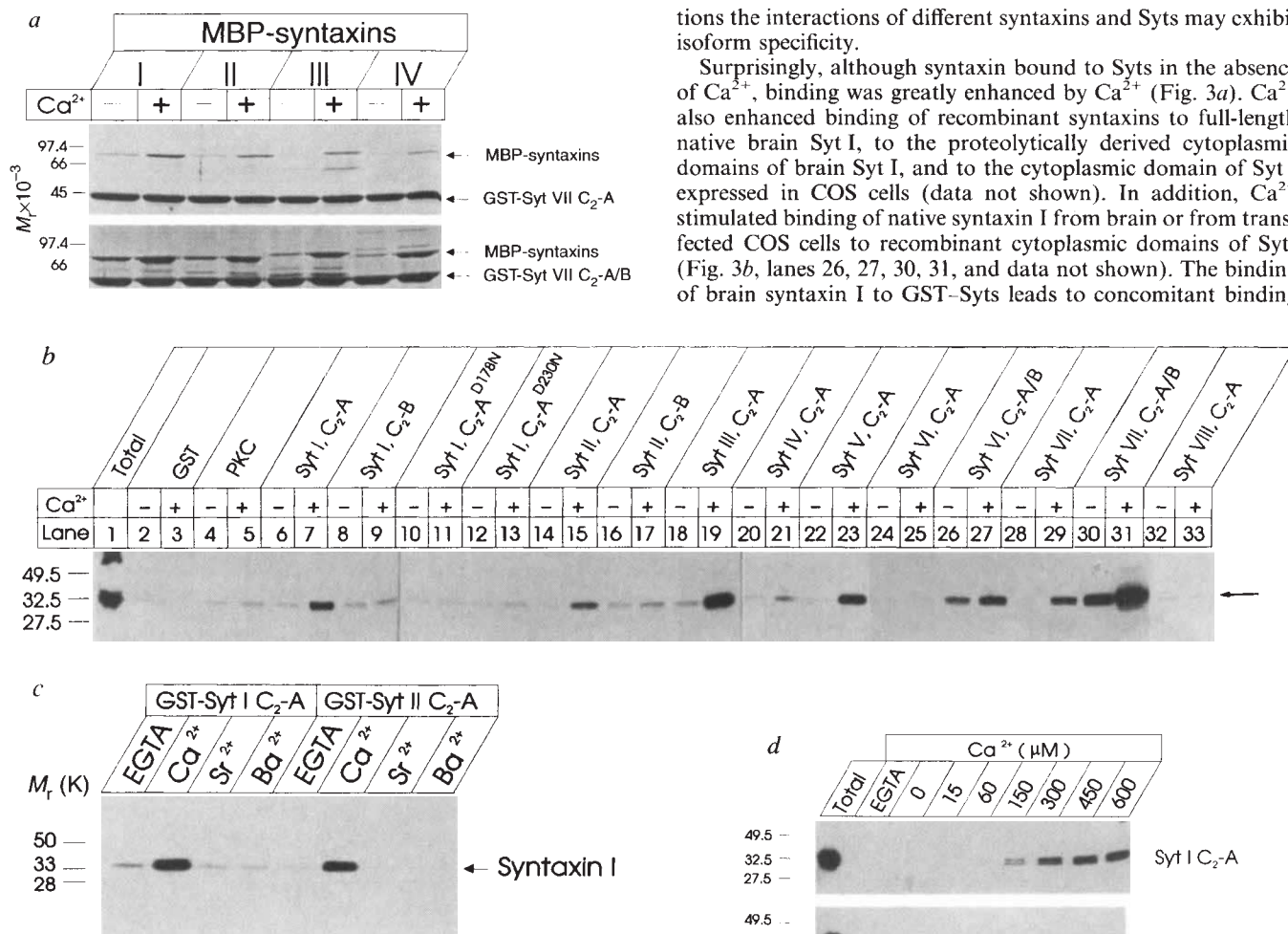


FIG. 3 a, Binding of syntaxins I–IV to Syt VII. Recombinant syntaxins fused to maltose-binding protein (MBP) were incubated with GST–Syt VII fusion proteins containing either the first C₂-domain (C₂-A) or both C₂-domains (C₂-A/B). Binding reactions were performed in the presence of 1 mM Mg²⁺ with or without 1 mM Ca²⁺ and analysed by SDS–PAGE and Coomassie-blue staining. Incubations of MBP–syntaxins without GST–Syts revealed no binding (not shown). b, Specificity of Ca²⁺-dependent interaction of syntaxin I with C₂ domains. Total brain extract (lane 1) was incubated with GST alone (lanes 2, 3) or with GST-fusion proteins containing the C₂ domain of protein kinase C (PKC; lanes 4, 5) or C₂ domains of synaptotagmins (lanes 6–33) attached to glutathione beads in the presence of 1 mM Mg²⁺ with or without 1 mM Ca²⁺. Proteins bound to washed beads were analysed by immunoblotting for syntaxin I (arrow) as shown, and for other synaptic proteins (not shown). C₂ domains analysed included the first (C₂-A) and second C₂ domain (C₂-B) of Syt I and II; mutants of the first C₂ domain of Syt I that do not bind phospholipid as a function of Ca²⁺ (C₂-A^{D178N} and C₂-A^{D230N})^{21,26}, the first C₂-domains of Syt III, IV, V and VIII, and the first and both C₂-domains of Syt VI and VII. M_r markers (in thousands) are indicated on the left. c, Ba²⁺ and Sr²⁺ do not activate syntaxin binding. GST–Syt I and II fusion proteins containing the first C₂ domains were incubated with brain extract containing the indicated cations at 1 mM and binding was measured by blotting for syntaxin I. d, Ca²⁺ affinity for syntaxin binding to the first C₂ domains of Syt I, II, III, V and VII. Syntaxin binding was measured as for b, with the indicated additions of Ca²⁺ in the presence of 1 mM Mg²⁺. For Syt III and VII, Ca²⁺ contaminating the zero-Ca²⁺ buffer was sufficient to trigger binding and an EGTA point was included to control for the Ca²⁺ contamination.

METHODS. Syntaxins I–IV were cloned into the filled-in *Eco*RI and *Hind*III sites of pMAL-c2 (New England Biolabs) and expressed and purified according to the manufacturer's protocol. To study interactions between recombinant proteins, purified MBP–syntaxins were incubated with GST–Syts attached to glutathione–agarose beads, or GST–syntaxins I–IV (ref. 30) were incubated with cytoplasmic domains of Syt I cleaved from GST–Syt fusion proteins by thrombin. To study binding of native

tions the interactions of different syntaxins and Syts may exhibit isoform specificity.

Surprisingly, although syntaxin bound to Syts in the absence of Ca^{2+} , binding was greatly enhanced by Ca^{2+} (Fig. 3a). Ca^{2+} also enhanced binding of recombinant syntaxins to full-length native brain Syt I, to the proteolytically derived cytoplasmic domains of brain Syt I, and to the cytoplasmic domain of Syt I expressed in COS cells (data not shown). In addition, Ca^{2+} stimulated binding of native syntaxin I from brain or from transfected COS cells to recombinant cytoplasmic domains of Syts (Fig. 3b, lanes 26, 27, 30, 31, and data not shown). The binding of brain syntaxin I to GST-Syts leads to concomitant binding

syntaxin I and Syt I, rat brains were homogenized in 4 mM HEPES–NaOH, pH 7.4, 1 mM EGTA, centrifuged at 100,000g, and washed in 10 mM HEPES–NaOH, pH 7.4, 1 M NaCl. Final membrane pellet was solubilized in buffer A (10 mM HEPES, 1 mM EGTA, 0.15 M NaCl, 2 mM $MgCl_2$, 1% Triton X-100), and insoluble proteins were removed by centrifugation. COS cells transfected with a syntaxin 1A expression vector were directly solubilized in buffer A. All buffers contained proteinase inhibitors (0.5 $mg\ l^{-1}$ pepstatin, 0.5 $mg\ l^{-1}$ aprotinin and 1 mM PMSF). Solubilized brain (~0.25 mg protein) or COS cell proteins were incubated with the indicated GST-fusion proteins (~20 μg protein) in 0.1 ml for more than 4 h at 4 °C in buffer A containing either an additional 5 mM EGTA or 3 mM Ca^{2+} . For the Ca^{2+} titration experiments, incubations were performed in Chelex-extracted buffer A without EGTA after addition of 6 mM EGTA, no addition, or addition of Ca^{2+} at the indicated concentrations. Beads were washed 3 times in the incubation buffers and bound proteins were analysed by SDS–PAGE and Coomassie-blue staining or immunoblotting using monoclonal antibodies against syntaxin I or Syt I, or monoclonal and polyclonal antibodies against SNAP-25, synaptobrevin, Na^+ / K^+ -ATPase, synaptophysin, synaptotagmin, and the NMDA receptor (not shown).

of the other components of the core complex, synaptobrevin and SNAP-25, but not of various control proteins (data not shown). These data reveal that the interactions of Syt I, II, III, V and VII with syntaxins are Ca^{2+} regulated (summarized in Table 1).

In Syt I, the first C_2 -domain is responsible for Ca^{2+} -dependent phospholipid binding¹¹. Tests for syntaxin binding to the first C_2 -domains of Syt I to VIII demonstrated that only those C_2 domains that bound phospholipids as a function of Ca^{2+} also bound syntaxin I in a Ca^{2+} -regulated manner (Fig. 3b). Mutations in these domains which abolished phospholipid binding also abolished syntaxin binding (lanes 10–13, Fig. 3b), indicating that the same amino acids are important for both. By contrast, a recombinant C_2 domain from protein kinase C (PKC), which exhibits Ca^{2+} -dependent phospholipid binding (B. Davletov and T.C.S., unpublished results) did not bind syntaxin (Fig. 3b, lanes 4 and 5). Ba^{2+} and Sr^{2+} , which trigger phospholipid binding to single C_2 domains¹¹, did not support syntaxin binding (Fig. 3c).

Thus single C_2 -domains of Syts have two independent Ca^{2+} -dependent activities, phospholipid binding and syntaxin binding. The location of these binding sites in the compact β -pleated structure of the C_2 domain²⁶ is unclear. The fact that the same Syt I mutations abolish both interactions indicates either that the responsible sites overlap, or that the Ca^{2+} -dependent conformational change leading to phospholipid binding is a prerequisite for a second Ca^{2+} -dependent change leading to syntaxin binding. The inability of Sr^{2+} to substitute for Ca^{2+} correlates well with its inability to activate the fast component of neurotransmitter release that is impaired in Syt-I-knockout mice^{3,27}.

The Ca^{2+} concentration dependence of syntaxin binding differs between Syt isoforms. Binding to the first C_2 domains of Syt I, II and V required more than 200 μM Ca^{2+} , whereas binding to Syt III and VII occurred in the low μM Ca^{2+} range (Fig. 3d, and data not shown). The syntaxin binding properties of these isoforms correlate with their relative sequence similarities (Fig. 1a). Therefore syntaxin binding by different Syts exhibits distinct Ca^{2+} regulation indicative of roles in different fusion reactions (summarized in Table 1).

Different Syts thus exhibit similar AP2 binding properties but distinct Ca^{2+} -dependent interactions with phospholipids and syntaxins. The interactions of the various Syts with neural and non-neural syntaxins are consistent with a general role in exocytosis, whereas their AP2 binding suggests a ubiquitous role in endocytosis. Their syntaxin and phospholipid binding properties divide Syts into three classes (summarized in Table 1): Ca^{2+} -independent Syts (Syt IV, VI and VIII), Syts that bind phospholipids and syntaxins at low Ca^{2+} concentrations (Syt III and VII), and Syts that bind phospholipids at low Ca^{2+} concentrations and syntaxins at high Ca^{2+} concentrations (Syt I, II and V). The Ca^{2+} dependence of syntaxin binding by the last class mirrors that of synaptic vesicle exocytosis¹² and indicates that fast neurotransmitter release may be triggered by the effect of high Ca^{2+} concentrations on Syt I, II or V, a hypothesis consistent with the phenotype of mice lacking Syt I³. An analogous role for the ubiquitous Syts in other membrane fusion reactions is supported by their interaction with non-neural syntaxins at much lower calcium levels. □

Received 27 January; accepted 28 April 1995.

- Bennett, M. K. & Scheller, R. H. A. *Rev. Biochem.* **63**, 63–100 (1994).
- Jahn, R. & Südhof, T. C. A. *Rev. Neurosci.* **17**, 219–246 (1994).
- Geppert, M. et al. *Cell* **79**, 717–727 (1994).
- Zhang, J. Z., Davletov, B. A., Südhof, T. C. & Anderson, R. G. W. *Cell* **78**, 751–760 (1994).
- Ullrich, U. et al. *Neuron* **13**, 1281–1291 (1994).
- Bommert, K. et al. *Nature* **363**, 163–165 (1993).
- Littleton, J. T., Stern, M., Schulze, K., Perlin, M. & Bellen, H. J. *Cell* **74**, 1125–1134 (1993).
- DiAntonio, A. & Schwarz, T. L. *Neuron* **12**, 909–920 (1994).
- Nonet, M. L., Grundahl, K., Meyer, B. J. & Rand, J. B. *Cell* **73**, 1291–1305 (1993).
- Brose, N., Petrenko, A. G., Südhof, T. C. & Jahn, R. *Science* **256**, 1021–1025 (1992).
- Davletov, B. A. & Südhof, T. C. *J. biol. Chem.* **268**, 26386–26390 (1993).
- Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. *Nature* **371**, 513–515 (1994).
- Perin, M. S., Fried, V. A., Mignery, G. A., Jahn, R. & Südhof, T. C. *Nature* **345**, 260–261 (1990).
- Geppert, M., Archer, B. T. & Südhof, T. C. *J. biol. Chem.* **266**, 13548–13552 (1991).
- Mizuta, M. et al. *J. biol. Chem.* **269**, 11675–11678 (1994).

- Hilbush, B. S. & Morgan, J. I. *Proc. natn. Acad. U.S.A.* **91**, 8195–8199 (1994).
- Hata, Y., Davletov, B., Petrenko, A. G., Jahn, R. & Südhof, T. C. *Neuron* **10**, 307–315 (1993).
- Perin, M. S. *J. biol. Chem.* **266**, 623–629 (1994).
- Bennett, M. K., Calakos, N. & Scheller, R. H. *Science* **257**, 255–259 (1992).
- Yoshida, A. et al. *J. biol. Chem.* **267**, 25925–25928 (1992).
- Davletov, B. A. & Südhof, T. C. *J. biol. Chem.* **269**, 28547–28550 (1994).
- Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. *Cell* **75**, 409–418 (1993).
- Hayashi, T. et al. *EMBO J.* **13**, 5051–5061 (1994).
- McMahon, H. et al. *Nature* **364**, 346–349 (1993).
- Bennett, M. et al. *Cell* **74**, 863–873 (1993).
- Sutton, R. B., Davletov, B. A., Berghuis, A. M., Südhof, T. C. & Sprang, S. R. *Cell* **80**, 929–938 (1995).
- Goda, Y. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12942–12946 (1994).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).
- Guan, K. L. & Dixon, J. E. *Analyt. Biochem.* **192**, 262–267 (1991).
- Hata, Y., Slaughter, C. A. & Südhof, T. C. *Nature* **366**, 347–351 (1993).

ACKNOWLEDGEMENTS. We thank T. Rosahl for his contribution to this study; R. Jahn, F. M. Brodsky and M. S. Robinson for antibodies; B. Davletov for advice; A.-Y. Chen for initial construction of pGEX plasmids; I. Leznicki, E. Borowicz and A. Roth for technical assistance; and M. S. Brown and J. L. Goldstein for advice. This study was partly supported by a grant from the Perot Family Foundation and by a fellowship grant from the Deutsche Forschungsgemeinschaft (to B.U.).

An excitatory amino-acid transporter with properties of a ligand-gated chloride channel

W. A. Fairman^{*†}, R. J. Vandenberg^{*††}, J. L. Arriza[†], M. P. Kavanaugh[†] & S. G. Amara^{*†§}

^{*} Howard Hughes Medical Institute and [†] Vollum Institute for Advanced Biomedical Research, Oregon Health Sciences University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97201, USA

EXCITATORY amino-acid transporters (EAATs) in the central nervous system maintain extracellular glutamate concentrations below excitotoxic levels and may limit the activation of glutamate receptors. Here we report the cloning of a novel human aspartate/glutamate transporter, EAAT4, which is expressed predominantly in the cerebellum. The transport activity encoded by EAAT4 has high apparent affinity for L-aspartate and L-glutamate, and has a pharmacological profile consistent with previously described cerebellar transport activities. In *Xenopus* oocytes expressing EAAT4, L-aspartate and L-glutamate elicited a current predominantly carried by chloride ions. This chloride conductance was not blocked by components that block endogenous oocyte chloride channels. Thus EAAT4 combines the re-uptake of neurotransmitter with a mechanism for increasing chloride permeability, both of which could regulate excitatory neurotransmission.

The properties of several recently cloned excitatory amino-acid transporters (EAAT1–3)¹ do not account for all pharmacologically distinguishable glutamate uptake activities in the cerebellum^{2–4}. To identify additional members of the EAAT family expressed in cerebellum, degenerate oligonucleotides complementary to two regions of amino-acid sequence similarity between the carriers were used to amplify complementary DNAs from human cerebellar messenger RNA. Two independent full-length cDNAs encoding EAAT4 were identified using a probe from one of these polymerase chain reaction (PCR) products. The amino-acid sequence predicted by the EAAT4 cDNAs exhibits 65%, 41% and 48% amino-acid identity to the human glutamate transporters EAAT1, 2, and 3 (ref. 1), respectively, and does not correspond to any glutamate transporter subtype isolated to date^{5–7} (Fig. 1a). Northern blotting analysis (Fig. 1b) of RNAs from both human brain and peripheral regions

[†] Present address: Department of Pharmacology, University of Sydney, Australia.

[§] To whom correspondence should be addressed.

demonstrates that the EAAT4 cDNA hybridizes to a single 2.4-kilobase (kb) mRNA in cerebellum and placenta. Although EAAT4 appears predominantly cerebellar by Northern blot analysis, more sensitive PCR-based assays indicate the EAAT4 is expressed, albeit at much lower levels, in brain stem, cortex and hippocampus.

Oocytes injected with EAAT4 cRNA exhibited ~30-fold higher saturable, sodium-dependent uptake of [3 H]L-aspartate and [3 H]L-glutamate than uninjected oocytes (Table 1). The properties of EAAT4 were also examined by analysis of excitatory amino-acid-induced currents in *Xenopus* oocytes voltage clamped at -60 mV. Application of L-aspartate or L-glutamate induced dose-dependent saturable inward currents in oocytes expressing EAAT4, but not in uninjected oocytes. The K_m values derived from current measurements were similar to those for radiolabelled flux (Table 1); these are ~10-fold lower than for other EAATs expressed in the *Xenopus* oocyte system¹. Structural analogues of glutamate, including L-trans-2,4-pyrrolidine dicarboxylic acid (PDC), L-quisqualate, and L- α -amino adipate, also elicited currents in oocytes expressing EAAT4 (Table 1). Kainic acid, which blocks glutamate transport in some brain regions²⁻⁴, neither elicited a current nor blocked an L-aspartate-

induced current (data not shown) at concentrations as high as 5 mM. The pharmacological properties of EAAT4, in particular L- α -amino adipate sensitivity and kainic acid insensitivity, are similar to those previously ascribed to uptake activities in the cerebellum²⁻⁴, and are consistent with the predominantly cerebellar localization of EAAT4 mRNA.

Excitatory amino-acid-induced currents observed in voltage-clamped oocytes expressing other members of this gene family exhibit strong inward rectification and do not reverse below +40 mV^{8,9,10}. In contrast, application of 10 μ M L-aspartate to oocytes expressing EAAT4 activated an outward current at potentials more positive than -20 mV (Fig. 2a). The voltage dependence of the L-aspartate-induced current was nearly linear over the range -100 mV to +40 mV, and reversed at -22 ± 1.6 mV ($n=18$; Fig. 2b). Although the equilibrium potential, E_{Cl} , may differ slightly in our oocytes, this reversal potential is close to previously reported values of E_{Cl} in *Xenopus* oocytes (-24 mV)¹¹. The reversal potential of the currents elicited by L-aspartate shifted by 57 ± 1.5 mV ($n=7$) per 10-fold change in $[Cl^-]_o$ (Fig. 2b, c), in accord with predictions of the Nernst equation for a chloride-selective conductance. For currents elicited with L-glutamate, the reversal potential shifted by 50 ± 2 mV

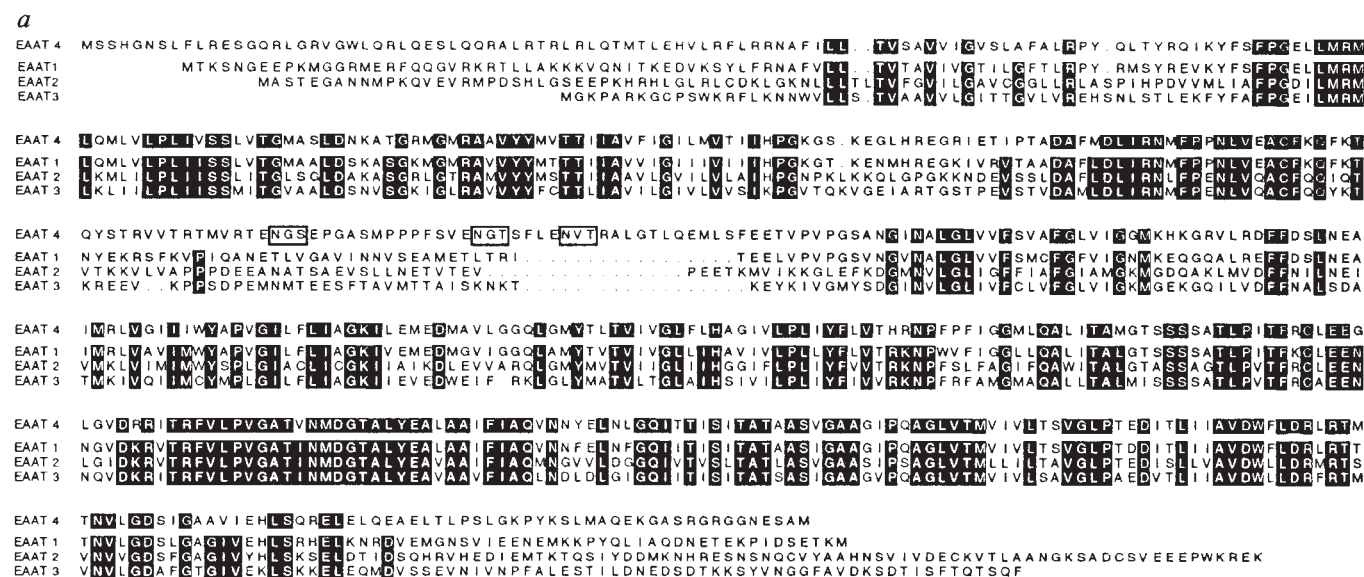
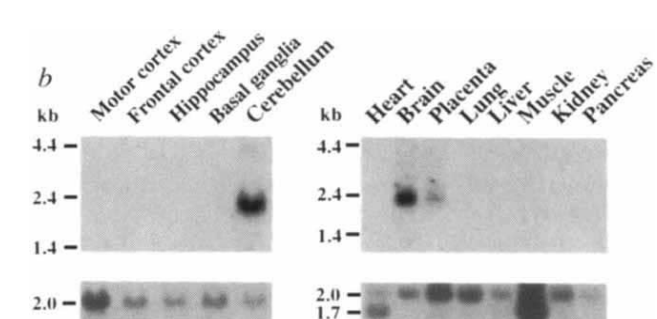


FIG. 1. Alignment of the amino-acid sequence of EAAT4 with human EAAT1, EAAT2 and EAAT3. a, The amino-acid sequences deduced from EAAT4 cDNA sequences are shown. Sequence identities between EAAT1-3 and EAAT1-4 are shaded in black. Potential sites of N-linked glycosylation in EAAT4 are boxed. b, Northern blotting analysis of EAAT4 RNA. An EAAT4-specific probe was hybridized to RNA from CNS (left) and peripheral tissues (right). Each lane of the human brain blot contains 20 μ g total RNA (left), and each lane of the human peripheral tissue RNA blot contains 2 μ g poly(A)⁺ RNA (right). Control hybridization of each blot with the β -actin probe is shown underneath.

METHODS. Degenerate oligonucleotides complementary to highly conserved sequences in previously cloned human transporters (EAAT1-EAAT3) were synthesized and used to amplify human cerebellar cDNA to identify additional transporter cDNAs. Oligonucleotide sequences were 5'-C GCG GGT ACC AA(T/C) CT(C/G) GT(C/A/G) (G/C) A(G/A) GC(T/C) TG(T/C) TT(T/C)-3' and 5'-C GCG TCT AGA(T/C) TG(A/G/T) GC(A/G/T) AT(A/G) AA(A/G) A(T/C) (G/T/C) GC(A/G/T)GC-3', which correspond to amino acids 189 to 196 and 434 to 441 of EAAT4. Human cerebellum RNA was isolated as described¹. This RNA was used as template for cDNA synthesis using the Superscript Choice System (GIBCO). PCR products were generated from 10 ng cerebellar cDNA by 35 cycles of denaturation (94 °C for 1 min), annealing (47 °C for 1 min) and extension (72 °C for 2 min) with Vent Polymerase (NE Biolabs) and 3 mM MgCl₂. A novel clone similar to the previously identified glutamate transporters was identified and used to screen a human motor cortex



library¹. Two independent clones containing 2.2 and 2.3 kilobase pair cDNA inserts were obtained by *in vivo* excision (Stratagene) and sequenced on both strands with synthetic oligonucleotide primers by dideoxy sequencing with Sequenase 2.0 (U.S. Biochemical). The nucleotide sequence has been deposited in GenBank (accession number U18244). A 1.7 kb probe specific for EAAT4 was excised and radiolabelled with α - 32 P-dCTP (NEN) by random primed synthesis (Boehringer Mannheim). Blots of RNA from human peripheral tissues (Clontech) and CNS regions were prepared, hybridized and washed as previously described¹.

TABLE 1 Pharmacological properties of EAAT4 substrate flux and currents

Substrate	K_m (μM)	$V_{\max}$ (fmol s^{-1})
[^3H]-L-aspartate	0.97 ± 0.29	3.26 ± 0.06
[^3H]-L-glutamate	2.49 ± 0.87	5.72 ± 0.52
Substrate	K_m (μM)	$I_{\max}$
L-aspartate	1.84 ± 0.46	1.0
L-glutamate	3.3 ± 0.4	0.37 ± 0.07
D-aspartate	2.5 ± 0.3	1.0 ± 0.1
L- α -amino adipate	168 ± 11	0.75 ± 0.02
L-quisqualate	99 ± 11	0.47 ± 0.06
L-homocysteate	403 ± 72	0.83 ± 0.05
PDC	2.6 ± 0.4	0.41 ± 0.01
Kainic acid	>5000	undetectable

Data shown are the mean $\pm$ s.e.m. of 3 to 5 independent determinations. The EAAT4 cDNA coding region was amplified using synthetic oligonucleotides and subcloned into pOTV³⁰ for the generation of cRNA for expression in *Xenopus laevis* oocytes. Protocols for the preparation, injection and electrophysiology of *Xenopus* oocytes have been described previously^{1,30}. Oocytes were voltage clamped at -60 mV and continuously superfused with ND96 (96 mM NaCl, 2 mM KCl, 1.8 mM CaCl_2 , 1 mM MgCl_2 , 5 mM HEPES, pH 7.5) containing varying concentrations of compounds. Current and flux measurements were fitted to the Michaelis-Menten equation, $V/V_{\max}$ or $I/I_{\max} = [S]/(K_m + [S])$, where $V_{\max}$ is the maximal transport rate, $I_{\max}$ is the maximal current, $[S]$ the substrate concentration, and K_m is the Michaelis constant. Doses of L-aspartate or L-glutamate used in K_m determinations for both [^3H]-L-aspartate and [^3H]-L-glutamate uptake and amino-acid-induced currents were 0.1, 0.3, 1, 3, 10, 30 and 100 μM . The maximal current generated by other compounds relative to L-aspartate was estimated by applying a maximal dose of L-aspartate (100 μM) before and/or after dose-response determinations. No compounds tested induced currents in uninjected or water injected oocytes. Uptake of 10 μM [^3H]-L-aspartate was linear for at least 20 min at room temperature (data not shown). [^3H]-L-aspartate (15.5 Ci mmol^{-1} , NEN) and [^3H]-L-glutamate (17.8 Ci mmol^{-1} , NEN) transport was measured in ND96 for 15 min at room temperature and terminated with 4 washes of ice-cold buffer. Specific uptake is defined as the radiolabelled uptake at each concentration for EAAT4-expressing oocytes ($n > 4$) minus the uptake in uninjected oocytes. Oocytes were solubilized in 0.1% SDS before scintillation counting in 10 ml of ScintiVerse (Fisher). All compounds tested were obtained from Sigma except L-quisqualate and L-trans-2,4-pyrrolidine dicarboxylic acid (RBI).

($n = 5$, data not shown) per 10-fold change in $[\text{Cl}^-]_o$. Complete removal of external chloride by substitution with gluconate abolished the L-aspartate-induced outward currents seen at depolarized potentials (Fig. 2b). In oocytes voltage clamped at -60 mV, [^3H]-L-aspartate transport in the absence of chloride (4.7 ± 0.3 fmol s^{-1} , $n = 8$) was not significantly different from [^3H]-L-aspartate transport in the presence of chloride

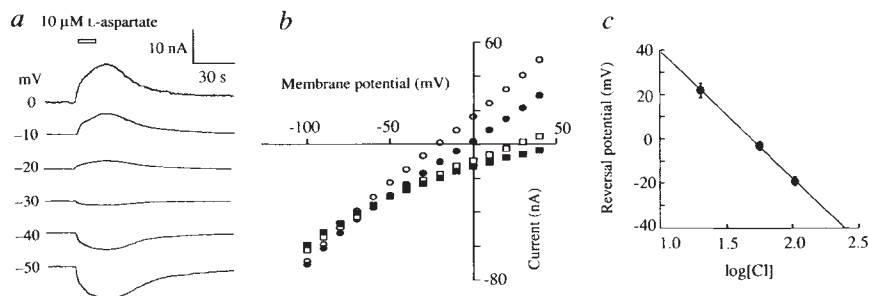
(5.9 ± 0.3 fmol s^{-1} , $n = 8$). Therefore [^3H]-L-aspartate transport is not absolutely dependent on external chloride. Substitution of sodium with choline completely and reversibly abolished L-aspartate-elicited currents ($n = 6$, Fig. 3a). These results are consistent with the sodium-dependent activation of a chloride conductance by excitatory amino acids.

The possibility that EAAT4 couples with a chloride channel endogenous to *Xenopus* oocytes was investigated. *Xenopus* oocytes express high levels of Ca^{2+} -activated chloride channels¹¹⁻¹³. Removal of external Ca^{2+} in the presence of 5 mM EGTA did not shift the reversal potential or significantly inhibit the amplitude of substrate-evoked currents ($94 \pm 6\%$ of control at -60 mV, $n = 6$). In addition, oocytes injected with BAPTA and clamped at -60 mV had L-aspartate-induced currents similar in amplitude (32 ± 5 nA, $n = 6$) to oocytes that were not injected with BAPTA (37 ± 3 nA, $n = 8$) (Fig. 3b). The same concentration of BAPTA completely abolished chloride currents elicited by the application of Ca^{2+} after treatment with the Ca^{2+} ionophore A23187 (Fig. 3b). Two chloride channel blockers, niflumic acid and NPPB (5-nitro-2-(3-phenylpropylamino)-benzoic acid) have been reported to inhibit peak Ca^{2+} -activated chloride currents in oocytes with K_i values of 17 μM ¹⁴ and 22 μM ¹⁵, respectively. Niflumic acid (100 μM) or NPPB (50 μM) did not inhibit L-aspartate-elicited currents, but increased the current amplitude at -60 mV by $86 \pm 12\%$ ($n = 5$), and $103 \pm 16\%$ ($n = 5$), respectively. Furthermore, L-aspartate-induced currents at -60 mV were not significantly inhibited by the non-selective chloride blockers, DIDS^{16,17} (4,4'-diisothiocyanostilbene-2,2'-disulphonate, 500 μM , $93 \pm 9\%$ of control, $n = 6$), and SITS¹⁷ (4-acetamido-4'-isothiocyanostilbene-2,2'-disulphonic acid, 500 μM , $87 \pm 7\%$ of control, $n = 3$). In oocytes, 50 μM DIDS has been shown to completely block endogenous Ca^{2+} -activated chloride channels¹⁶. These results suggest that endogenous chloride channels do not mediate the L-aspartate-induced chloride conductance in oocytes expressing EAAT4.

To investigate the relationship of chloride flux to L-aspartate translocation in oocytes expressing EAAT4, [^3H]-L-aspartate (3 μM) was applied for 10 min to individual oocytes held at either -60 mV (below E_{Cl}) or -5 mV (above E_{Cl}). Upon removal of [^3H]-L-aspartate, currents returned to baseline and remained constant as the oocytes were washed in ND96 for 2 min at the same holding potential. Specific uptake by EAAT4-expressing oocytes was more than 10-fold greater than uninjected oocytes. At -60 mV, an inward current of 11.2 ± 0.7 nA was elicited with an inward flux of 4.1 ± 0.2 fmol s^{-1} ($n = 5$), whereas at -5 mV an outward current of 3.9 ± 0.8 nA and an inward flux of 3.8 ± 0.2 fmol s^{-1} ($n = 4$) were observed. Thus L-aspartate transport occurs regardless of the direction of chloride flux, and does not appear significantly affected by membrane potential over this voltage range. Although substrate flux has

FIG. 2 a, Steady-state currents in a representative oocyte expressing EAAT4 at the indicated membrane potentials induced by superfusion of 10 μM L-aspartate for the time indicated by bar. b, Chloride dependence of L-aspartate-elicited current-voltage relation. Steady-state currents recorded during 50 ms voltage jumps in ND96 (see Table 1) were subtracted from those in the presence of 10 μM L-aspartate in ND96 containing either 104 mM (open circles), 56 mM (filled circles), 20 mM (open squares) or 0 mM (filled squares) chloride replaced by equimolar gluconate. The I-V curves of a representative cell from 7 independent determinations are shown. c, Chloride dependence of the reversal potential. The slope for chloride dependence of reversal potential for L-aspartate-elicited currents is -57 ± 1.5 mV per decade. Data points represent the mean $\pm$ s.e.m. ($n = 7$).

METHODS. Oocytes were prepared as described in Table 1. Data collection and analysis used pCLAMP software (Axon Instruments). Chloride



substitution experiments entailed substituting equimolar concentrations of gluconate for chloride to maintain osmolality. A 3 M KCl agar bridge from the recording chamber to a 4 M KCl reservoir containing a Ag/AgCl electrode was used to avoid offset voltages associated with changes in chloride concentration.

been observed to be voltage dependent for EAAT2 (ref. 10), serotonin transport has been reported to be relatively independent of voltage¹⁹. Studies of other glutamate transporters suggest that glutamate influx is coupled to the cotransport of two sodium ions and the countertransport of potassium¹⁹ and hydroxyl ions²⁰ to generate a net inward movement of one positive charge per transport cycle. Assuming one charge per cycle for EAAT4, an inward flux of $4.1 \pm 0.2 \text{ fmol s}^{-1}$ ($n=5$) would generate a current of 0.4 nA, rather than the 11.2 nA current observed. This discrepancy between the observed and expected current suggests that the chloride conductance represents ~95% of the L-aspartate evoked current at -60 mV.

Substrate-dependent differences in the current-voltage relation were apparent in oocytes expressing EAAT4. Increasing concentrations of L-aspartate (Fig. 3c) or L-glutamate (data not shown) resulted in saturable increases in the membrane conductance without affecting the reversal potential for each substrate. However, the reversal potentials of the currents varied with the substrates. The reversal potential was $-22 \pm 1.4 \text{ mV}$ ($n=18$) for the L-aspartate elicited current, $-3.5 \pm 0.4 \text{ mV}$ ($n=3$) for L-glutamate (Fig. 3d) and $-6.0 \pm 0.7 \text{ mV}$ ($n=3$) for L- α -amino adipate (data not shown). Substrate-specific differences were also observed in the chloride dependence of the reversal potential of excitatory amino-acid-induced currents. The shift observed with L-aspartate ($57 \pm 1.5 \text{ mV}$) fits more closely than L-glutamate ($50 \pm 2 \text{ mV}$) to predictions of the Nernst equation for a chloride-selective conductance. In addition, there appears to be an inverse relation for I_{max} and V_{max} between L-aspartate and L-glutamate (Table 1). The observed substrate-dependent reversal potentials

reflect the combined contributions of the chloride conductance and the current directly attributable to the cotransport of substrates (determined by the V_{max} of transport and the ionic stoichiometry). Variations in the relative contribution of each component are a likely explanation for the observed differences in E_{rev} , because the L-glutamate-dependent current mediated by EAAT4 appears to reflect a greater transport-associated component and a lesser chloride component than that induced by L-aspartate. Excitatory amino-acid induced currents mediated by other members of this transporter family have relatively small chloride components and reverse at correspondingly positive potentials²¹. Although both components of the current share similar requirements for activation, our data support a model in which the amino-acid flux is not directly coupled to chloride flux.

Chloride-dependent glutamate transport activities in synaptosomes and other brain preparations have been reported²²⁻²⁴, but these studies have not demonstrated that chloride ions are cotransported with substrates. Glutamate-activated, sodium-dependent chloride currents in the cone cells of the salamander retina have been ascribed to a glutamate receptor-gated chloride channel²⁵ or a protein with properties of both a glutamate transporter and an ion channel²⁶. The conductance of EAAT4 has several properties in common with that of the cone cell: it is rapidly activated by glutamate, is sodium dependent, and is relatively selective for chloride.

Several distinct classes of proteins which function as chloride channels have been identified, but there seem to be no unifying structural motifs, which would imply a common mechanism for

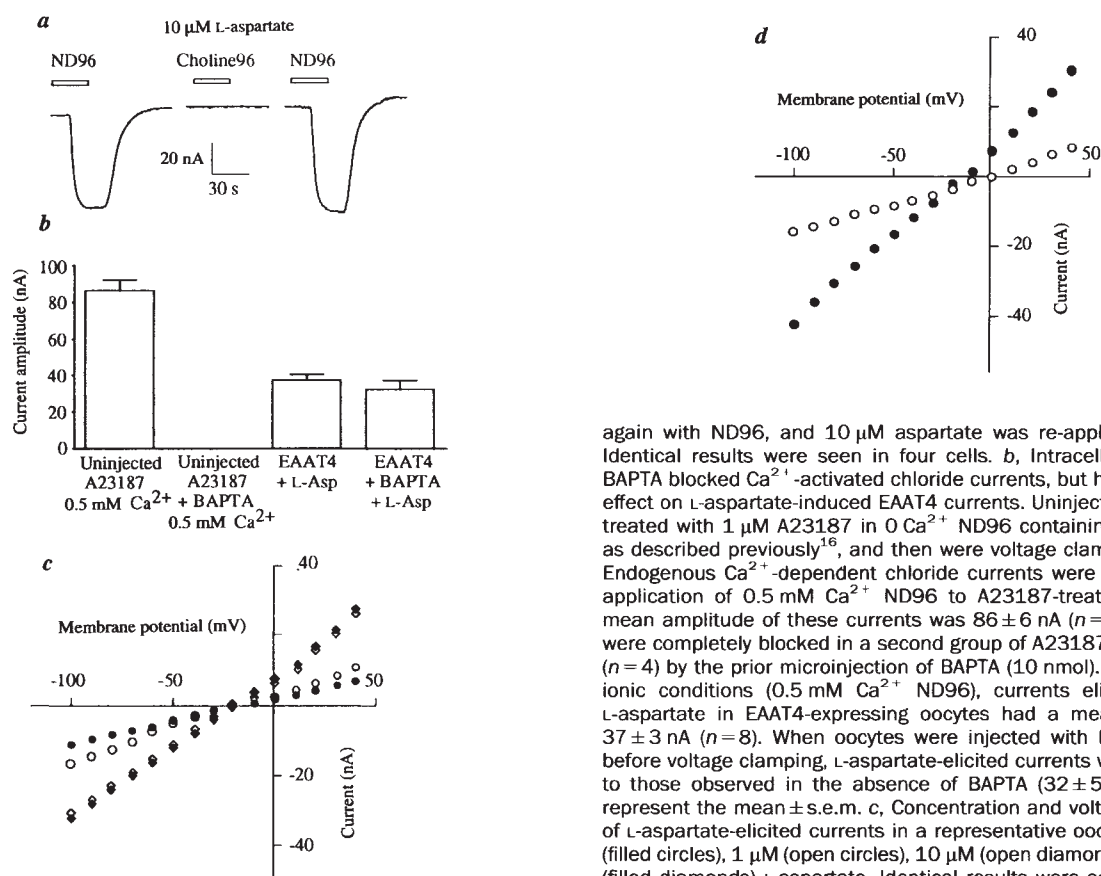


FIG. 3 **a**, Sodium dependence of L-aspartate-induced currents. The left trace shows the response to 10 μM L-aspartate in ND96. The same oocyte was then bathed in a sodium-free solution (choline96) and 10 μM aspartate was applied (middle trace). The oocyte was perfused

again with ND96, and 10 μM aspartate was re-applied (right trace). Identical results were seen in four cells. **b**, Intracellular injection of BAPTA blocked Ca^{2+} -activated chloride currents, but had no significant effect on L-aspartate-induced EAAT4 currents. Uninjected oocytes were treated with 1 μM A23187 in 0 Ca^{2+} ND96 containing 0.5 mM EGTA, as described previously¹⁶, and then were voltage clamped at -60 mV. Endogenous Ca^{2+} -dependent chloride currents were activated by the application of 0.5 mM Ca^{2+} ND96 to A23187-treated oocytes. The mean amplitude of these currents was $86 \pm 6 \text{ nA}$ ($n=5$). The currents were completely blocked in a second group of A23187-treated oocytes ($n=4$) by the prior microinjection of BAPTA (10 nmol). Under the same ionic conditions (0.5 mM Ca^{2+} ND96), currents elicited by 10 μM L-aspartate in EAAT4-expressing oocytes had a mean amplitude of $37 \pm 3 \text{ nA}$ ($n=8$). When oocytes were injected with BAPTA (10 nmol) before voltage clamping, L-aspartate-elicited currents were comparable to those observed in the absence of BAPTA ($32 \pm 5 \text{ nA}$, $n=6$). Bars represent the mean $\pm$ s.e.m. **c**, Concentration and voltage dependence of L-aspartate-elicited currents in a representative oocyte, with 0.3 μM (filled circles), 1 μM (open circles), 10 μM (open diamonds), and 100 μM (filled diamonds) L-aspartate. Identical results were seen in four cells. **d**, Current-voltage relation obtained for currents elicited by L-aspartate and L-glutamate. I-V curves were generated with the application of a maximal dose (100 μM) of L-aspartate (filled circles) or L-glutamate (open circles) in a representative oocyte expressing EAAT4. Identical results were seen in three cells.

chloride movement. These families include the voltage-gated chloride channels, and the ligand-gated chloride channels (GABA and glycine receptors). Members of the ATP-binding cassette (ABC) transporter family, such as the cystic fibrosis transmembrane conductance regulator (CFTR) and the multidrug resistance pump (MDR1), have also been proposed to function as chloride channels^{27,28}. Mutational analyses of the CFTR protein²⁹ have identified residues which contribute to ionic selectivity, indicating that a channel-like pore is integral to the CFTR protein. EAAT4 appears to be an additional example of a membrane protein combining structural and functional aspects of both a transporter and an ion channel. EAAT4 with its dual functions as a transporter and a glutamate-gated chloride channel could have significant consequences for synaptic transmission. Its transport function would reduce the amount of neurotransmitter available for activating postsynaptic receptors, and its capacity for enhancing chloride permeability could alter neuronal excitability. □

Received 16 January; accepted 28 April 1995.

1. Arriza, J. L. et al. *J. Neurosci.* **14**, 5559–5569 (1994).
2. Coyle, J. T. & Ferkany, J. J. *Neurosci. Res.* **16**, 491–503 (1985).
3. Fletcher, E. J. & Johnston, G. A. R. *J. Neurochem.* **57**, 911–915 (1991).
4. Robinson, M. B., Sinor, J. D., Dowd, L. A. & Kerwin, J. F. Jr *J. Neurochem.* **60**, 167–179 (1993).

5. Kanai, Y. & Hediger, M. A. *Nature* **360**, 467–471 (1992).
6. Storck, T., Schulte, S., Hofmann, K. & Stoffel, W. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10955–10959 (1992).
7. Pines, G. et al. *Nature* **360**, 464–467 (1992).
8. Kanai, Y. et al. *J. Biol. Chem.* **269**, 20599–20606 (1994).
9. Klöckner, U., Storck, T., Conradt, M. & Stoffel, W. *J. Biol. Chem.* **268**, 14594–14596 (1993).
10. Wadiche, J. L., Arriza, J. L., Amara, S. G. & Kavanaugh, M. P. *Neuron* **14**, 1019–1027 (1995).
11. Barish, M. E. *J. Physiol., Lond.* **342**, 309–325 (1983).
12. Milei, R. & Parker, I. J. *J. Physiol., Lond.* **357**, 173–183 (1984).
13. Milei, R. *Proc. R. Soc. B215*, 491–497 (1982).
14. White, M. M. & Aylwin, M. *Molec. Pharmacol.* **37**, 720–724 (1990).
15. Wi, G. & Hamill, O. P. *Pflügers Arch.* **420**, 227–229 (1992).
16. Cunningham, S. A., Worrell, R. T., Benos, D. J. & Frizzel, R. A. *Am. J. Physiol.* **262**, 783–788 (1992).
17. Greger, R. *Meth. Enzym.* **191**, 793–810 (1990).
18. Mager, S. et al. *Neuron* **12**, 845–859 (1994).
19. Barbour, B., Brew, H. & Attwell, D. *J. Physiol., Lond.* **436**, 169–193 (1991).
20. Bouvier, M., Szatkowski, M., Amato, A. & Attwell, D. *Nature* **360**, 471–474 (1992).
21. Wadiche, J. L., Amara, S. G. & Kavanaugh, M. P. *Neuron* (in the press).
22. Kuhar, M. J. & Zarbin, M. A. *J. Neurochem.* **31**, 251–256 (1978).
23. Zaczek, R. et al. *Neuropharmacology* **26**, 281–287 (1987).
24. Pin, J.-P., Bockaert, J. & Reasens, M. *FEBS Lett.* **175**, 31–36 (1984).
25. Sarantis, M., Everett, K. & Attwell, D. *Nature* **332**, 451–453 (1988).
26. Eliasof, S. & Werblin, F. J. *Neurosci.* **13**, 402–411 (1993).
27. Bear, C. E. et al. *Cell* **68**, 809–818 (1992).
28. Valverde, M. A. et al. *Nature* **355**, 830–833 (1992).
29. Anderson, M. P. et al. *Science* **253**, 202–205 (1991).
30. Arriza, J. L. et al. *J. Biol. Chem.* **268**, 15329–15332 (1993).

ACKNOWLEDGEMENTS. We thank G. Murdoch, S. Eliasof, M. Sonders and C. Jahr for critical review of this manuscript; S. Eliasof and J. Wadiche for advice and constructive criticisms; and W. Zhang for preparation and injection of *Xenopus* oocytes. This work was supported by the Howard Hughes Medical Institute, a Hitchings Award from the Burroughs Wellcome Fund, and NIH grants (S.G.A. and M.P.K.). R.J.V. was supported by a fellowship from the Human Frontiers of Science Program.

Antigen presentation mediated by recycling of surface HLA-DR molecules

Valérie Pinet^{*†}, Marco Vergelli[‡], Roland Martin[‡], Oddmund Bakke[§] & Eric O. Long^{*}

^{*} Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

[‡] Neuroimmunology Branch, National Institute of Neurological Diseases and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

[§] Department of Biology, University of Oslo, N-0316 Oslo, Norway

CLASS II histocompatibility molecules associate with peptides derived from antigens that are processed in endocytic compartments. Antigen presentation to class II-restricted T cells generally requires newly synthesized class II molecules¹, associated invariant chain^{2,3}, and HLA-DM^{4,5}. Exceptions to these rules have been reported^{6–10}, but without description of an underlying mechanism. Here we show that presentation of immunodominant epitopes in the haemagglutinin protein of influenza virus and in myelin basic protein correlates with recycling of surface HLA-DR molecules. Truncation of either one of the α or β cytoplasmic tails virtually eliminated internalization of HLA-DR molecules and presentation of haemagglutinin from inactive virus particles. In contrast, the invariant chain-dependent presentation of matrix antigen from the same virus particles was unaffected by these truncations. Thus HLA-DR cytoplasmic tails are not required for the conventional presentation pathway, but jointly contribute a signal for an alternative pathway involving internalization of HLA-DR molecules.

Recycling of cell-surface class II molecules, although still controversial¹¹, has been suggested as a pathway for presentation of some antigens^{8,12,13}. A sensitive assay was developed using biotinylated F(ab) fragments to measure internalization and

recycling of class II and other molecules. As expected¹⁴, the transferrin receptor (TfR) was rapidly internalized from the surface of a B-cell line and levelled off with 20% of the molecules remaining at the cell surface (Fig. 1a). In contrast, more than 90% of HLA-A molecules remained at the cell surface during the chase at 37 °C (not shown). HLA-DR molecules internalized rapidly but levelled off with ~80% of the molecules remaining at the cell surface (Fig. 1b). Surface HLA-DR molecules associated with invariant chain (Ii) cannot account for the observed internalization, because such $\alpha\beta$ Ii complexes represent only 3% of the total surface pool of HLA-DR¹⁵. Furthermore, transfected human fibroblasts expressing wild-type DR molecules, alone or with Ii, internalized DR molecules actively, with only ~50% of the molecules remaining at the cell surface after 1 hour (Fig. 1c). As reported for other fibroblasts¹⁶, this Ii-independent internalization results in class II molecules being clearly visible in vesicular structures, detected by immunofluorescence with permeabilized cells (not shown).

Rapid recycling of internalized molecules back to the cell surface accounts for the observed plateau of molecules remaining at the cell surface in internalization assays. Conversely, the observed fraction of molecules returning to the surface will level off owing to rapid re-internalization. Our recycling data (Fig. 1, insets) are consistent with this prediction with plateaux at ~45% for TfR (Fig. 1a) and ~80% for HLA-DR (Fig. 1b), reached within 10–15 min. The higher plateau for HLA-DR reflects the lesser internalization and larger relative size of the extracellular pool, as compared with TfR. Recycling of internalized DR molecules was also efficient in the transfected fibroblasts (Fig. 1c). Thus our data demonstrate extensive and rapid internalization and recycling of HLA-DR molecules on a human B-cell line in agreement with results using a different assay¹⁷, and on transfected human fibroblasts.

HLA-DR molecules with truncated cytoplasmic tails, expressed in transfected fibroblasts, failed to internalize (Fig. 1c). As expected from the lack of internalization, vesicular staining for intracellular DR molecules was not observed in cells expressing truncated DR molecules in the absence of Ii (not shown). Surface levels of DR molecules were virtually identical on transfected cells expressing Ii with wild-type DR molecules (8G9-DR) or truncated DR molecules (8G9- $\Delta\alpha\Delta\beta$) (Fig. 2a),

[†] Present address: Laboratoire d'Immunologie, Hôpital Saint-Eloi, INSERM U291, 34295 Montpellier Cedex 5, France.

providing an opportunity to test how internalization and recycling affect antigen presentation. Cells were pulsed with ultra-violet-inactivated influenza virus particles and tested for lysis by a haemagglutinin H3-specific or a matrix M1-specific T-cell clone (Fig. 2b). Whereas H3 presentation occurred in 8G9-DR cells, it was clearly reduced in 8G9- $\Delta\alpha\Delta\beta$ cells. In contrast, M1 presentation was essentially identical in 8G9-DR and 8G9- $\Delta\alpha\Delta\beta$ cells. Averages from three independent experiments (Fig. 2c) clearly establish that H3 presentation, but not M1 presentation, correlated with internalization and recycling of mature $\alpha\beta$ dimers.

The ability of 8G9- $\Delta\alpha\Delta\beta$ cells to process virus particles and to present the M1 epitope to class II-restricted T cells demonstrates that the last 10 and 12 residues of the DR α and β cytoplasmic tails, respectively, are dispensable for the proper transport of II-associated class II molecules to antigen-processing compartments, for the peptide loading step, and for their transport from such compartments to the cell surface. Therefore the lack of H3 presentation by 8G9- $\Delta\alpha\Delta\beta$ cells is not due to a generalized defect in virus uptake or in triggering T cells. Differences in T-cell sensitivity or peptide affinity cannot account for the lack of H3 presentation, because H3-specific T cells are about 5 times more sensitive than the M1-specific T cells⁸, and HLA-DR1 has a higher affinity for H3 than for M1 peptide¹⁸. The synthetic H3 peptide was presented to H3-specific T cells as

effectively by 8G9- $\Delta\alpha\Delta\beta$ cells as 8G9-DR cells, thus eliminating the possibility that the DR cytoplasmic tails are uniquely required for stimulation of those H3-specific T cells. We therefore conclude that recycling of DR molecules is used for H3 presentation. Previous work reported that autoreactive T cells, but not antigen-specific T cells, required class II cytoplasmic tails for a proper response^{19,20}. Although attributed to a defect in class II intracellular signalling, this lack of T-cell stimulation may relate to the alternative presentation pathway described here.

To test for the role of each cytoplasmic tail, other transfected fibroblasts were generated, expressing a normal α -chain with a truncated β -chain, or a normal β -chain with a truncated α -chain. These 'hemi-truncated' DR molecules were expressed at surface levels comparable to those on 8G9- $\Delta\alpha\Delta\beta$ and 8G9-DR cells (data not shown). Surprisingly, truncation of either one of the cytoplasmic tails (Fig. 3a) reduced internalization of DR molecules to about 5% (Fig. 3b). Presentation of H3 antigen, but not M1 antigen, was drastically reduced (Fig. 3c), confirming our conclusion that recycling of surface DR molecules is required for presentation of the H3 antigen. Consistent with this, efficient H3 presentation by B cells pulsed with virus particles was more rapid than M1 presentation (Fig. 4). The requirement for both cytoplasmic tails, neither of which carries an obvious internalization signal, suggests that the two cytoplasmic tails

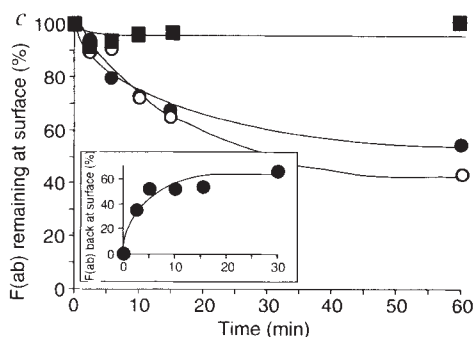
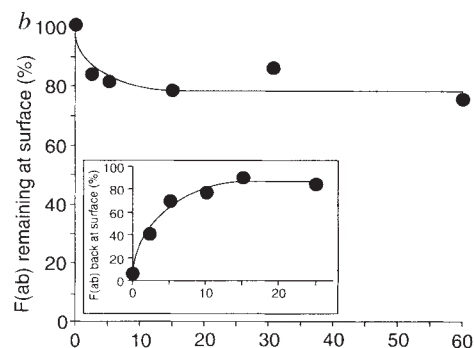
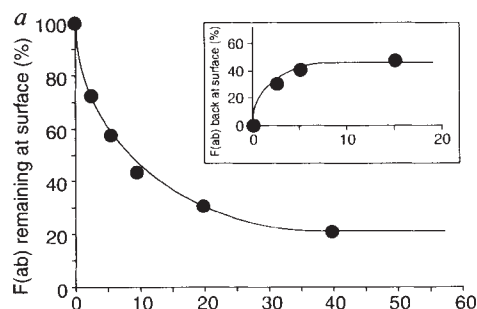


FIG. 1 Truncation of the cytoplasmic tails of DR α - and β -chains abrogates internalization of cell-surface DR molecules. The extent of internalization was deduced from the amount of biotinylated F(ab) fragments remaining at the cell surface at 37 °C, as quantified by [¹²⁵I]-streptavidin binding. To measure recycling of internalized molecules, cells were treated with glutathione on ice to remove surface-associated biotin, and recultured at 37 °C (insets). F(ab) fragments were specific for Tfr (a) and HLA-DR (b, c). a, b, B-cell line 721.45, and c, transfected human fibroblasts RSV-DR (filled circles), 8G9-DR (open circles) and 8G9- $\Delta\alpha\Delta\beta$ (filled squares).

METHODS. Complementary DNAs encoding the DR α - and β -chains in the vector CDM8 (ref. 27) were mutagenized by PCR to introduce stop codons that removed the last 10 amino acids of the α -chain and the last 12 of the β -chain. The remaining amino acids in the cytoplasmic tails are KGVRK (for α) and RNQK (for β). These cDNAs were cotransfected with pSV2-gpt into the human fibroblast clone 8G9, expressing both p33 and p35 forms of li, and the control transfectant RSV expressing no li, as described²⁷. Transfected fibroblasts RSV-DR and 8G9-DR have been described²⁷. The antibodies used were the IgG1 anti-Tfr IG12 (ref. 28), anti-HLA-A PA2.1 (ATCC clone HB 117), and anti-class II β -chain XD5.A11 (ref. 29). F(ab) fragments prepared by standard papain digestion and HPLC purification were biotinylated with NHS-SS-biotin (Pierce Chemicals) for 2 h at 4 °C. Free NHS-SS-biotin was removed by 3 consecutive centrifugations through a Centricon-10. An ELISA test with specific anti-F(ab) or anti-Fc antibodies (Sigma) was used to demonstrate purity of the F(ab) fragments. Media used during internalization and recycling experiments were RPMI-1640 with 1% bovine serum albumin (BSA) for the B-cell line 721.45, and Dulbecco's modified Eagle's medium (DMEM) with 1% BSA for human fibroblasts. For internalization experiments, cells were incubated for 2 h at 4 °C at 15×10^6 cells ml^{-1} in the presence of 10–30 $\mu\text{g ml}^{-1}$ F(ab)-biotin. After washing, cells were resuspended in warm medium at 15×10^6 cells

ml^{-1} and chased at 37 °C. Portions (150 μl , 2×10^6 cells) were taken at various times and internalization of the molecules stopped by adding cold medium. Cells were resuspended in 150 μl DPBS with 1% BSA containing 0.3 μCi [¹²⁵I]-streptavidin. After 30 min at 4 °C, free [¹²⁵I]-streptavidin was removed by centrifugation through a mixture of dinonyl phthalate:dibutyl phthalate (Kodak), 1:3 for fibroblasts, and 1:6 for B cells. Pellets were counted in a gamma counter. For recycling experiments, cells were allowed to internalize surface molecules tagged with F(ab)-biotin for 1 h, incubated twice for 10 min at 4 °C in 50 mM glutathione, 75 mM NaCl, 10 mM EDTA, 75 mM NaOH, 1% BSA to remove surface-bound biotin. After washing, cells were resuspended in warm medium at 15×10^6 cells ml^{-1} . Portions (150 μl , 2×10^6 cells) were taken at various times, and recycling of the molecules was stopped by adding cold medium. The fraction of recycled molecules was calculated relative to the molecules internalized in 1 h.

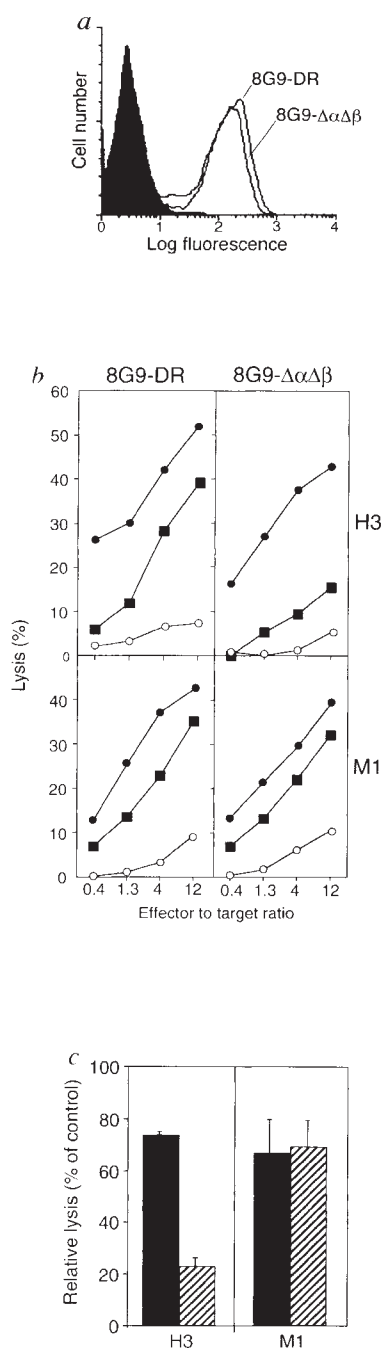


FIG. 2 Presentation of the H3 antigen but not the M1 antigen requires the cytoplasmic tails of DR molecules. **a**, Cell-surface expression of DR molecules on transfected fibroblasts 8G9-DR and 8G9- $\Delta\alpha\Delta\beta$ assayed by flow cytometry with the monoclonal antibody L243 (ATCC clone HB 55). The filled profiles represent cells incubated with FITC-conjugated goat anti-mouse Ig alone. **b**, Transfected fibroblasts were incubated with antigen for 4 h and tested for lysis by DR1-restricted T-cell clones exactly as described⁹. Upper: lysis by the H3-specific clone E1.9 of cells either untreated (open circles), pulsed with $10 \mu\text{g ml}^{-1}$ synthetic peptide H3 307–318 (filled circles), or with a 10^{-2} dilution of ultraviolet-inactivated influenza virus⁸ (filled squares). Lower: lysis by the M1-specific clone C3.5 of the same cells, except that the synthetic peptide was M1 18–29. Left: 8G9-DR cells expressing li with wild-type DR1 molecules. Right: 8G9- $\Delta\alpha\Delta\beta$ cells expressing li with truncated DR1 molecules. **c**, Average from three experiments performed as in (b) and expressed in relative lysis at an effector to target ratio of 4 (per cent of specific lysis divided by the percentage of lysis with a saturating amount of peptide). Filled bars, 8G9-DR cells; hatched bars, 8G9- $\Delta\alpha\Delta\beta$ cells. Left: lysis by the H3-specific clone E1.9. Right: lysis by the M1-specific clone C3.5.

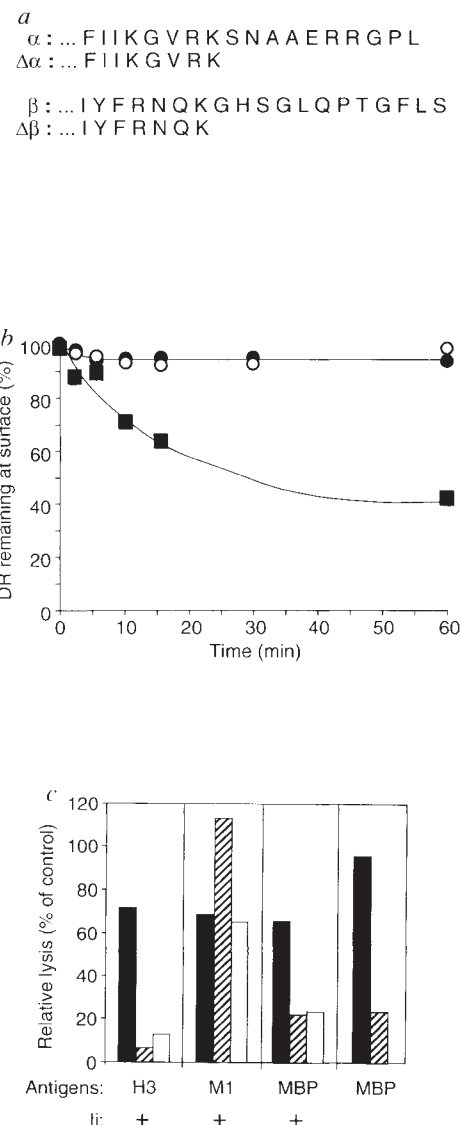


FIG. 3 Internalization of DR molecules and presentation of H3 and MBP antigens require both α and β cytoplasmic tails. **a**, Amino-acid sequence of the cytoplasmic tails of α -, $\Delta\alpha$ -, β -, $\Delta\beta$ -chains of DR molecules (in one-letter code). **b**, DR molecules remaining at the cell surface were quantified as in Fig. 1 for 8G9-DR cells (filled squares), 8G9- $\beta\Delta\alpha$ (filled circles) and 8G9- $\alpha\Delta\beta$ (open circles). **c**, Presentation of the indicated antigens (H3 and M1 from ultraviolet-inactivated influenza virus; MBP in lipid-bound form) by human fibroblasts expressing either full-length α - and β -chains of HLA-DR (filled bars), or full-length β and truncated α (hatched bars), or full-length α and truncated β (white bars) in the presence or absence of li, as indicated. **METHODS.** Experiments with ultraviolet-inactivated influenza virus were performed as in Fig. 2. Specific lysis with synthetic peptide was 39%, 38% and 40% (for H3), and 31%, 17% and 31% (for M1), with 8G9-DR, 8G9- $\beta\Delta\alpha$ and 8G9- $\alpha\Delta\beta$ cells, respectively. To avoid trypsinization after the pulse with MBP, fibroblasts were seeded at 2000 cells per well in U-bottomed 96-well plates and labelled for 15 h with sodium [^{51}Cr]chromate ($2.5 \mu\text{Ci}$ in $50 \mu\text{l}$). After two washes, the adherent fibroblasts were pulsed for 2 h in serum-free medium containing $15 \mu\text{g ml}^{-1}$ of synthetic MBP peptide 151–170 or $45 \mu\text{g ml}^{-1}$ of a preparation of intact, lipid-bound MBP²¹. Similar data were obtained with $15 \mu\text{g ml}^{-1}$ MBP. After a wash, DR1-restricted MBP-specific T cells, generated as described³⁰, were added at different effector to target ratios. The assay was completed as in Fig. 2. Specific lysis with synthetic peptide 151–170 was 40%, 40%, 35%, 32% and 37% for 8G9-DR, 8G9- $\beta\Delta\alpha$, 8G9- $\alpha\Delta\beta$, RSV-DR, and RSV- $\beta\Delta\alpha$ cells, respectively. Data are presented as relative lysis at an effector to target ratio of 30, as described in Fig. 2.

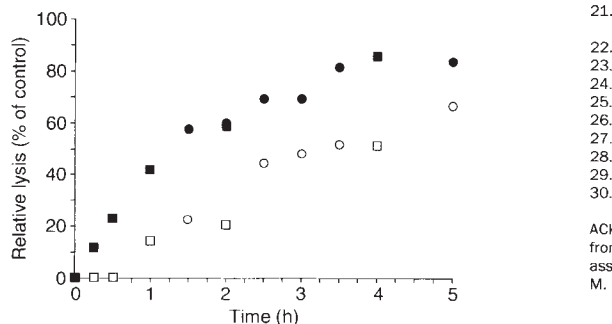


FIG. 4 Kinetics of H3 and M1 presentation. The human B-cell line 721.45 was pulsed for the indicated times with a 10^{-5} dilution of ultra-violet-inactivated influenza virus⁸ and tested for lysis by the H3-specific clone E1.9 (filled symbols) and by the M1-specific clone C3.5 (open symbols) at an effector to target ratio of 8. Squares and circles represent data obtained in two separate experiments. Data are expressed as relative lysis as described in Fig. 2. Specific lysis with synthetic peptide was 81.5% and 78.2% for H3, 55.3% and 53.9% for M1, in the two experiments, respectively.

may interact jointly with other molecules that control internalization.

To test whether the presentation pathway involving internalization of DR molecules can be used by other antigens, cells were pulsed with myelin basic protein (MBP) that was purified in its native lipid-bound form²¹, and were assayed for lysis by HLA-DR1-restricted MBP-specific T cells (Fig. 3c). Presentation was greatly reduced in 8G9- $\alpha\Delta\beta$ and 8G9- $\beta\Delta\alpha$ cells, even though these two cells presented a synthetic MBP peptide as efficiently as 8G9-DR cells. H3 (ref. 8) and, as reported previously²², MBP presentation did not require Ii (Fig. 3c). The requirement for internalization was also observed in the absence of Ii (Fig. 3c). Thus the new presentation pathway described here applies to different types of immunologically relevant antigen.

Complexes of class II molecules with high-affinity peptides do not detectably dissociate^{23,24}. However, a fraction of class II molecules is occupied by lower affinity peptides²⁵ and by Ii-derived CLIP peptides²⁶ that may exchange with other peptides after internalization of class II molecules. The compartment for peptide loading on recycling class II molecules is probably distinct from the one where newly synthesized class II molecules are transported, because H3 failed to be presented by cells expressing Ii with truncated DR molecules, in which M1 presentation was normal. The pathway described here may be useful for the presentation of antigens that are rapidly degraded after uptake by antigen-presenting cells. □

Received 28 February; accepted 18 April 1995.

- Davidson, H. W., Reid, P. A., Lanzavecchia, A. & Watts, C. *Cell* **67**, 105–116 (1991).
- Viville, S. et al. *Cell* **72**, 635–648 (1993).
- Bikoff, E. K. et al. *J. exp. Med.* **177**, 1699–1712 (1993).
- Morris, P. et al. *Nature* **368**, 551–554 (1994).
- Fling, S. P., Arp, B. & Plous, D. *Nature* **368**, 554–558 (1994).
- Kakiuchi, T., Watanabe, M., Hozumi, N. & Nariuchi, H. *J. Immunol.* **145**, 1653–1658 (1990).
- Nadimi, F. et al. *Eur. J. Immunol.* **21**, 1255–1263 (1991).
- Pinet, V., Mainat, M. S. & Long, E. O. *J. Immunol.* **152**, 4852–4860 (1994).
- Brooks, A. G., Campbell, P. L., Reynolds, P., Gautam, A. M. & McCluskey, J. J. *Immunol.* **153**, 5382–5392 (1994).
- Stebbins, C. C., Loss, G. E. Jr, Elias, C. G., Chervonsky, A. & Sant, A. J. *J. exp. Med.* **181**, 223–234 (1995).
- Cresswell, P. A. *Rev. Immunol.* **12**, 259–293 (1994).
- Adorini, L., Appella, E., Doria, G., Cardinaux, F. & Nagy, Z. A. *Nature* **342**, 800–803 (1989).
- Harding, C. V. & Unanue, E. R. *J. Immunol.* **142**, 12–19 (1989).
- Klausner, R. D. et al. *J. Biol. Chem.* **258**, 4715–4724 (1983).
- Roche, P. A., Teletski, C. L., Stang, E., Bakke, O. & Long, E. O. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8581–8585 (1993).
- Simonsen, A., Momburg, F., Drexler, J., Hammerling, G. J. & Bakke, O. *Int. Immunol.* **5**, 903–917 (1993).
- Reid, P. A. & Watts, C. *Immunology* **77**, 539–542 (1992).
- Rothbard, J. B. & Geffer, M. L. A. *Rev. Immunol.* **9**, 527–565 (1991).
- Nabavi, N. et al. *J. Immunol.* **142**, 1444–1447 (1989).
- St-Pierre, Y., Nabavi, N., Ghogawala, Z., Glimcher, L. H. & Watts, T. H. *J. Immunol.* **143**, 808–812 (1989).

- Riccio, P., Bobba, A., Romito, E., Minetola, M. & Quagliariello, E. *Neuroreport* **5**, 689–692 (1994).
- Bodmer, H., Viville, S., Benoist, C. & Mathis, D. *Science* **263**, 1284–1286 (1994).
- Lanzavecchia, A., Reid, P. A. & Watts, C. *Nature* **357**, 249–252 (1992).
- Nelson, C. A., Petzold, S. J. & Unanue, E. R. *Nature* **371**, 250–252 (1994).
- Germain, R. N. & Rinker, A. G. Jr, *Nature* **363**, 725–728 (1993).
- Chicz, R. M. et al. *Nature* **358**, 764–768 (1992).
- Long, E. O., LaVaute, T., Pinet, V. & Jaraquemada, D. *J. Immunol.* **153**, 1487–1494 (1994).
- Davis, J. E. & Cresswell, P. *J. Immunol.* **144**, 990–997 (1990).
- Radka, S. F., Machamer, C. E. & Cresswell, P. *Hum. Immunol.* **10**, 177–186 (1984).
- Martin, R. et al. *J. Immunol.* **148**, 1359–1366 (1992).

ACKNOWLEDGEMENTS. We thank P. Riccio for a preparation of MBP (obtained with support from the Italian MS society); P. Roche and B. Dobberstein for advice; A. Ashford for technical assistance; E. Mellins for discussion; P. Cresswell for monoclonal antibodies; and R. Germain, M. Marks and P. Roche for comments on the manuscript.

Antigenic oscillations and shifting immunodominance in HIV-1 infections

Martin A. Nowak*, Robert M. May*, Rodney E. Phillips†, Sarah Rowland-Jones†, David G. Lalloo†, Steven McAdam†, Paul Klenerman†, Britta Köppe†, Karl Sigmund†, Charles R. M. Bangham† & Andrew J. McMichael†

* Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, UK

† Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

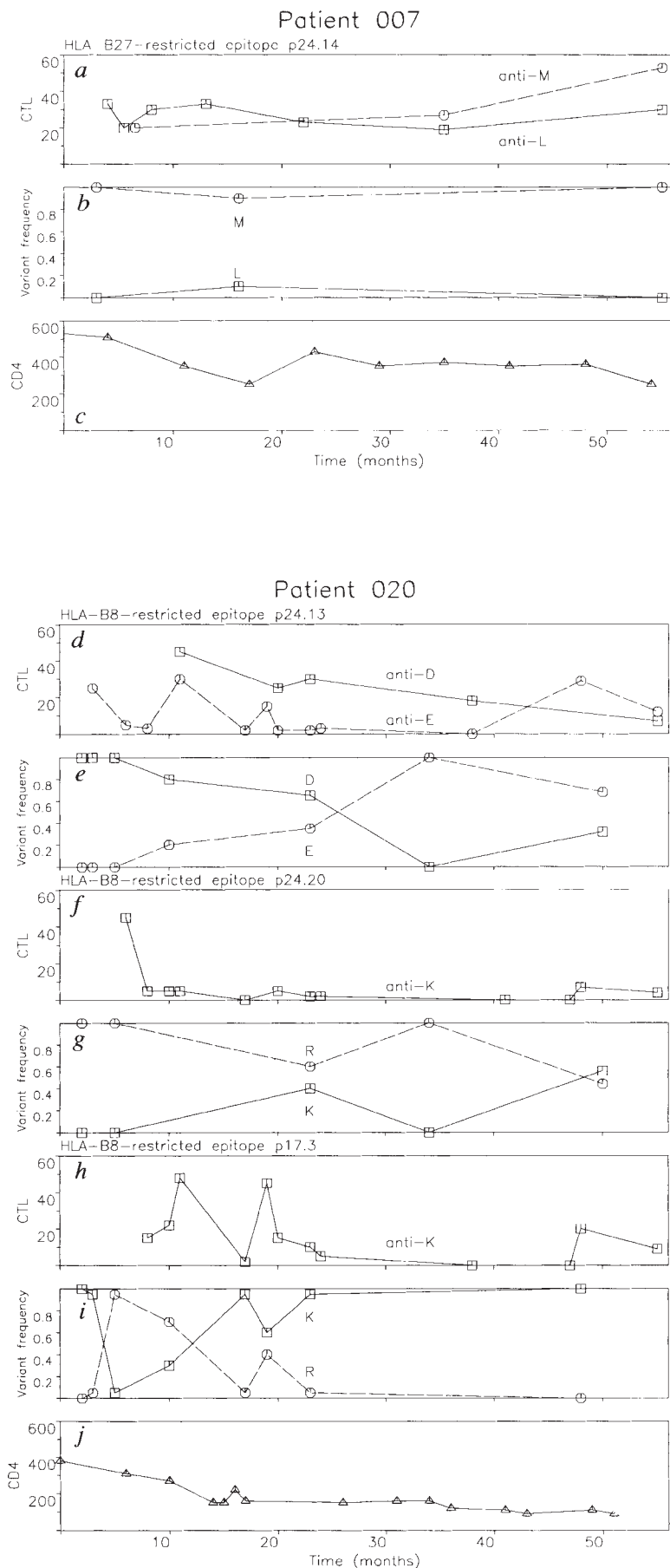
‡ Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria

A TYPICAL protein antigen contains several epitopes that can be recognized by cytotoxic T lymphocytes (CTL), but in a characteristic antiviral immune response *in vivo*, CTL recognize only a small number of these potential epitopes, sometimes only one^{1,2}, this phenomenon is known as immunodominance^{1–10}. Antigenic variation within CTL epitopes has been demonstrated for the human immunodeficiency virus HIV-1 (ref. 11) and other viruses^{12–17} and such 'antigenic escape' may be responsible for viral persistence. Here we develop a new mathematical model that deals with the interaction between CTL and multiple epitopes of a genetically variable pathogen, and show that the nonlinear competition among CTL responses against different epitopes can explain immunodominance. This model suggests that an antigenically homogeneous pathogen population tends to induce a dominant response against a single epitope, whereas a heterogeneous pathogen population can stimulate complicated fluctuating responses against multiple epitopes. Antigenic variation in the immunodominant epitope can shift responses to weaker epitopes and thereby reduce immunological control of the pathogen population. These ideas are consistent with detailed longitudinal studies of CTL responses in HIV-1 infected patients. For vaccine design, the model suggests that the major response should be directed against conserved epitopes even if they are subdominant.

In HIV-1 infection, there are strong CTL responses to several epitopes, and evidence is mounting that CTL are important in controlling the virus^{18–23}. HIV persistence and disease progression may rely partly on evasion of CTL responses by antigenic variation. We have followed CTL responses in two HIV-1 infected haemophiliac patients for 55 months (Fig. 1) and find that their responses are very different. Patient 007, who remains well, has a sustained CTL response to a single HLA-B27-restricted gag epitope. A weak response to pol was found on only one occasion and no CTL responses have been found to nef or env, nor to any other epitope in gag. In the gag epitope we found one predominant and one rare variant, but both

FIG. 1 Serial studies of CTL recognition of epitope variants in two HIV-1 positive haemophiliac patients. Patient 007 has a sustained response to the HLA-B27-restricted epitope gag p24.14(263–272). **a**, Percentage of specific lysis of HLA-match target cells treated with the peptides KRWIIMGNK (single-letter amino-acid codes, M variant) or KRWIILGNK (L variant). **b**, Sequencing of proviral DNA revealed that virtually all virus in peripheral blood mononuclear cells (PBMC) consists of the M variant rather than the L variant, which represents the database consensus sequence. **c**, Patient 007 is a slow progressor, with a slow decline in CD4 T-cell count to 300 cells per μ l. Patient 020 made responses to three HLA-B8-restricted epitopes in gag, designated p24.13(259–267), p24.20(329–337) and p17.3(24–32). **d**, The patient's CTL fluctuate in their capacity to recognize the two variants of epitope p24.13, GEIYKRWII (E variant) and GDIYKRWII (D variant), probably because of changes in the relative abundance of specific CTL clones. Both peptides bind HLA B8 with similar efficiency (S.McA. *et al.*, manuscript submitted). **e**, Changes in frequency of the p24.13 D and E variants. Initially only the D variant was detected in proviral DNA. Loss of the CTL response against the E variant at around 20 months was associated with its emergence as the only detectable variant. This appears to be an example of escape from a CTL response at a time when the circulating CTL clones fail to recognize the variant; later a crossreactive response returned and the selection pressure altered. **f**, CTL response to p24.20 variants. This response must be considered in conjunction with the response to p24.13, 60 amino acids upstream in HIV gag. The CTL response against the K variant was initially present and then lost. **g**, Frequency of p24.20 variants. The peptide DCKTILKAL (R variant), which cannot be recognized by the patient's CTL although it binds to HLA B8, (S. McA. *et al.*, submitted), predominated. Later, the recognizable variant, DCKTILKAL (K variant), dominated but there were no specific CTL detectable. The linkage of variants in p24.13 and p24.20 changed: at 23 months the D+R variants were linked, whereas at 34 months the E+R variants were the majority species. **h**, The CTL response against p17.3 was variable as the predominant sequence shifted from GGKKKYK (K variant) to GGRKKYK (R variant) and back. CTL never recognized the R variant¹¹, although it binds to HLA B8 (S.McA. *et al.*, submitted). **i**, Changing frequency of the p17.3 variants. The K variant became more abundant when the CTL response to this epitope was low. **j**, Serial CD4 counts. Patient 020 had a progressive loss of CD4 T cells, developing AIDS with a CD4 T-cell count of <200 per μ l, and pneumocystis pneumonia. He died of a gastrointestinal haemorrhage.

METHODS. CTL cultures were induced by incubating PBMC with autologous phytohaemagglutinin(PHA)-activated PBMC for 10–14 days (ref. 11, and S.McA. *et al.*, submitted). Killer-to-target ratios were between 30:1 and 50:1. Standard errors of the CTL responses were less than 10% of the specific lysis values. Peptides were synthesized by f-moc chemistry on an automated Zinsser peptide synthesizer. Proviral DNA was extracted from lymphocytes after brief (5–7 d) culturing with PHA and subjected to nested polymerase chain reactions (PCR)¹¹. At each time point we sequenced at least 20 M13 recombinant clones. Amplification was carried out with positive and negative controls. Sequence data from the adjacent epitopes p24.13 and p24.20 were obtained from a PCR reaction which amplified both epitopes simultaneously.



variants are seen by the patient's CTL. Patient 020, who progressed to AIDS in 8 years, showed variation within three HLA-B8-restricted gag epitopes¹¹. We found variants in all three epitopes that were not seen at certain time points by the patient's CTL. There is some evidence for selective increase of such escape variants: the p24.13E variant increased in frequency from 0 to 100% at a time when specific CTL failed to recognize this epitope. Similarly, a poor CTL response to the p17.3K variant is associated with accumulation of this variant. But the pattern is complex overall, and we observe changes in the frequency of variant viruses with shifting immunodominance in the CTL responses.

To explain these observations and to develop a general framework for understanding immunodominance and antigenic variation in several epitopes, we designed a new mathematical model for the dynamic interaction between virus mutants and activated CTL. Each virus mutant has several epitopes, and in each epitope there can be a number of variant peptide sequences inducing different lines of CTL. For simplicity, we discuss a model with two epitopes, but our results can be generalized to several epi-

topes. Virus dynamics can be represented as:

$$dv_{ij}/dt = v_{ij}(r_{ij} - p_i x_i - q_j y_j) \quad (1)$$

where $i = 1 \dots n_1$ and $j = 1 \dots n_2$; immune responses against epitope A as:

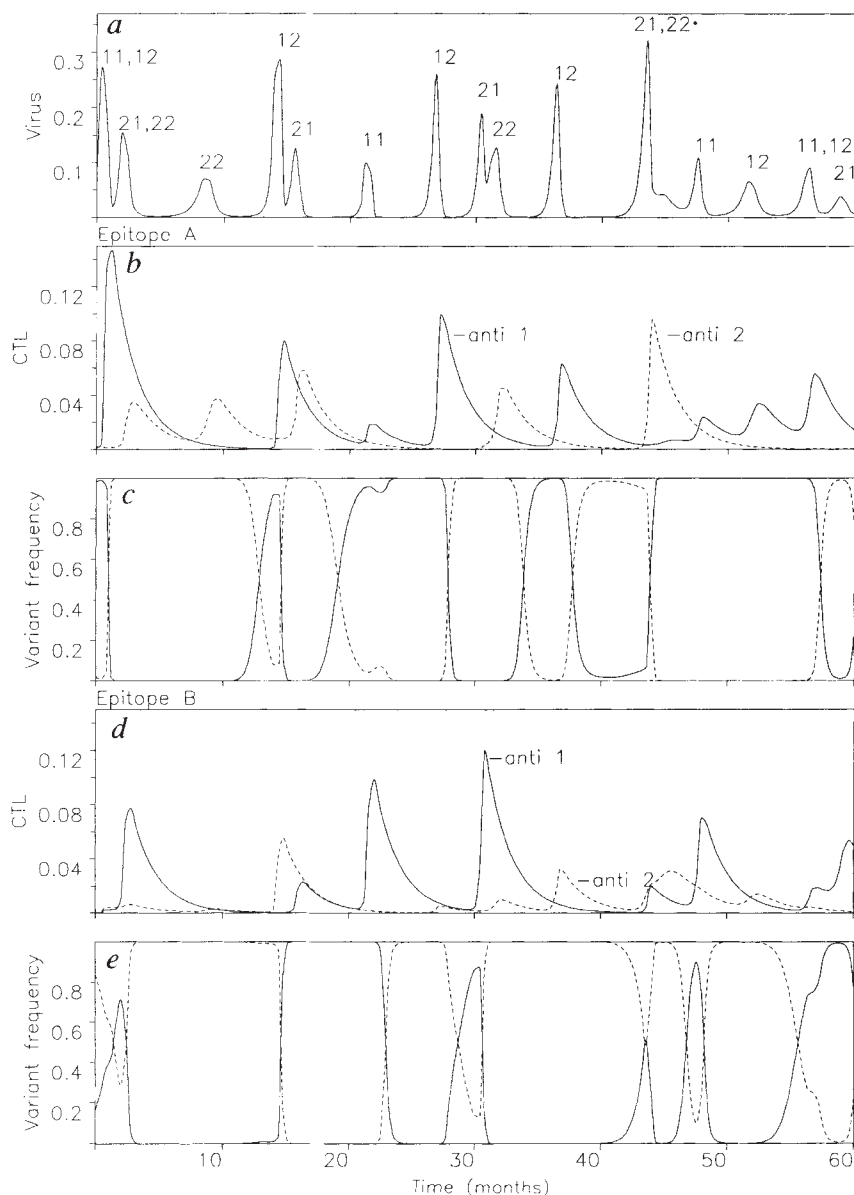
$$dx_i/dt = \eta c_i v_{i*} + x_i(c_i v_{i*} - b) \quad (2)$$

where $i = 1 \dots n_1$; and immune responses against epitope B as:

$$dy_j/dt = \eta k_j v_{*j} + y_j(k_j v_{*j} - b) \quad (3)$$

where $j = 1 \dots n_2$. Here v_{ij} denotes the abundance of virus variants with sequence i in epitope A and sequence j in epitope B. There are n_1 sequences for epitope A and n_2 for epitope B. The variables x_i and y_j denote CTL directed at sequence i of epitope A and sequence j of epitope B, respectively. The virus variant v_{ij} reproduces at the rate r_{ij} . Virus-infected cells are killed by CTL responses at the rates $p_i x_i v_{ij}$, where p_i and q_j are rate constants specifying the kinetics of removal of infected cells. CTL are stimulated by their specific epitope sequence (in association with HLA presentation). They are either produced by activation

FIG. 2 Antigenic oscillations and fluctuating immunodominance in a model with two epitopes. Peaks of viral abundance that consist of antigenically different variants are accompanied by oscillations in the size and specificity of the CTL responses. All virus variants are present at the beginning of the simulation; there are no additional mutational events. In each epitope we have two antigenically different variants that induce specific CTL responses. The 4 peptide sequences have different immunogenicities, and the 4 virus variants have different replication rates. The figure shows transient dynamics; the oscillations are damped on a time scale of $1/\eta$. a, Total virus abundance; b, CTL response against epitope A; c, frequency of the two different peptides in epitope A; d, CTL response against epitope B; e, frequency of the two different peptides in epitope B. The parameters of the simulation are: $r_{11} = 0.2$, $r_{12} = 0.15$, $r_{21} = 0.16$, $r_{22} = 0.1$, $c_1 = 1$, $c_2 = 1.1$, $k_1 = 1.9$, $k_2 = 0.8$, $p_i = q_j = 5$, $b = 0.02$, $\eta = 0.0001$.



from a pool of precursor cells (at the rates $\eta c_i v_{i*}$ and $\eta k_j v_{j*}$) or by proliferation of already activated cells (at the rates $c_i x_i v_{i*}$ and $k_j y_j v_{j*}$), where $v_{i*} = \sum_{j=1}^{n_2} v_{ij}$ and $v_{j*} = \sum_{i=1}^{n_1} v_{ij}$. Thus a particular CTL clone recognizes all viruses that have the specific sequence in the appropriate epitope; that is, x_i is directed at (and is stimulated by) v_{i*} , whereas y_j recognizes v_{j*} . The rate constants c_i and k_j describe the immunogenicities of sequence i in epitope A and sequence j in epitope B, respectively. The factor η describes the ratio of the rate at which CTL are activated from precursor cells over the rate of proliferation of already activated CTL. A small η implies that most activated CTL arise from proliferation of already activated cells. In the absence of antigenic stimulation, the activated CTL decline at the rates $b x_i$ and $b y_j$, where b represents the decay constant.

In our model, the immunogenicity of a peptide is defined as the rate at which it stimulates specific CTL. In molecular terms, several factors contribute to immunogenicity: a peptide is likely to be highly immunogenic if it binds the presenting HLA molecule with high affinity^{6,7}, if the resulting HLA-peptide complex is present in abundance on the surface of the infected cell, if this complex has a high affinity for the T-cell receptor, and if the reacting T-cells have high precursor frequency. The abundance of an epitope will be influenced by the concentration of the source protein, by its intrinsic stability and localization, and by the presence of appropriate proteolytic cleavage sites that lead to efficient antigen processing⁵.

An intrinsic feature of our model is competition among immune responses against different epitopes. Immunodominance arises naturally as a consequence of this competition. Consider a homogeneous virus population with two epitopes that have different immunogenicities and therefore induce CTL responses at different rates. Suppose that both CTL responses

have equal decay rates in the absence of antigenic stimulation. An equilibrium with a persistent virus population arises when the CTL have reduced the virus population to a level where CTL stimulation equals CTL decay. As the two epitopes have different immunogenicities, such an equilibrium is only possible for one of the two CTL responses. The same argument applies to situations with immune responses against many potential epitopes. In a homogeneous virus population there is competitive exclusion among the responses against different epitopes and eventually only a single response will persist. For a heterogeneous virus population, the situation is more complex²⁴ and antigenic variation generally leads to coexisting responses against several epitopes.

In situations with antigenic variation in multiple epitopes, the model generates complex oscillatory dynamics. There are distinct peaks in viral abundance, which are often dominated by single viral genotypes and occur whenever the responses against particular variants have declined to low levels because of a lack of stimulation (Fig. 2). The timescale of such oscillations is roughly determined by the time it takes for an activated CTL response to decline in the absence of further stimulation. For acute virus infections, activated CTL decline within 2–4 weeks after infection²⁵. For these 'antigenic oscillations' to occur, it is not essential for mutation continuously to generate new antigenic material. Antigenic oscillations are different from the previously described 'antigenic drift', where the emergence of a new variant gives rise to a peak in viral abundance²⁶. Antigenic oscillations arise as a consequence of the nonlinear dynamics of the immune responses acting on existing viral diversity.

A central point here is an understanding of the events following the emergence of a new mutant, in situations with several potential epitopes. All earlier experimental or theoretical studies of escape dynamics are essentially based, explicitly or implicitly, on models with single epitopes. In such models, immunological escape by an epitope confers a selective advantage, if the phenotype of the mutant is unaffected; such an escape mutant will increase in abundance and dominate the population until it induces a new specific response. If it fails to induce such a response, the escape mutant grows to fixation.

But antigenic variation in models with multiple epitopes is qualitatively different from the simple escape dynamics of single epitope models. We find situations where the escape mutant does not reach fixation or dominate the population. The immune system may not respond to the new peptide even if it is potentially immunogenic. The new mutant may induce a shift in immunodominance to another epitope, even if the homogeneous population of the mutant induces a response against the peptide in which the mutation occurred. In multiple epitope models, the most important consequence of antigenic variation is a shift of the immunodominant response to other (weaker) epitopes. This can increase the viral load and can thus represent a route to disease progression.

Consider a homogeneous virus population, v_{11} , which is exposed to CTL responses against two epitopes, A and B. The immunogenicity of epitope A is c_1 and of epitope B is k_1 . If $c_1 > k_1$, the response against epitope A is immunodominant. Suppose an escape mutant, v_{21} , emerges in epitope A. The immunogenicity of the variant sequence in epitope A is denoted by c_2 , and the replication rates of the original variant and the escape mutant by r_{11} and r_{21} , respectively. The CTL responses against peptides 1 and 2 of epitope A are denoted by x_1 and x_2 , the response against peptide 1 of epitope B by y_1 . For the following considerations we use equations (1–3) in the limiting case $\eta = 0$. The emergence of the escape mutant leads to one of four outcomes (Fig. 3). (1) It can simply lead to a diversification in epitope A without stimulating an immune response against epitope B. This happens if $1/c_1 + 1/c_2 < 1/k_1$. The system converges to oscillations around an equilibrium with $v_{11}, v_{21}, x_1, x_2 > 0$ and $y_1 = 0$. (2) The new mutant may not elicit a specific immune response to epitope A, but may induce a partial shift in immuno-

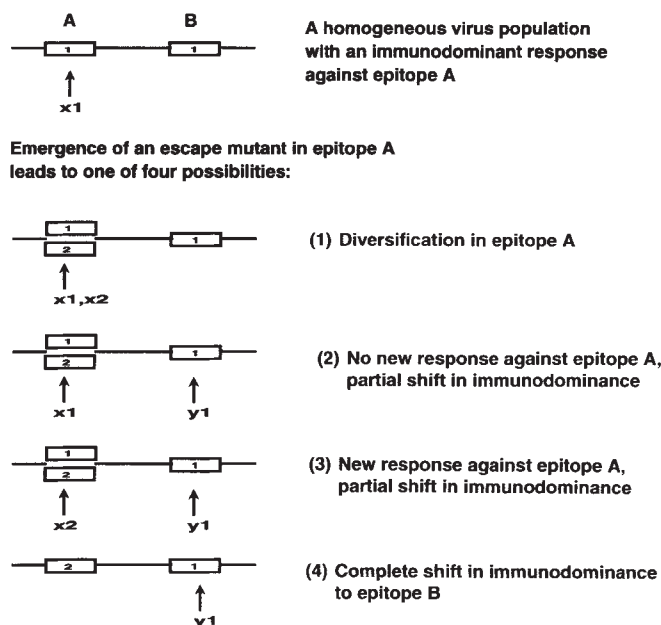


FIG. 3 Antigenic variation can shift immunodominance. The emergence of an escape variant in an immunodominant epitope, A, leads to one of 4 possible outcomes. (1) Diversification in epitope A, leading to a virus population with 2 epitope variants in A, both of which are seen by specific CTL. (2) Partial shift of immunodominance to epitope B, no specific response against the escape variant in A. (3) Partial shift of immunodominance to epitope B, together with a change in specificity of the response to epitope A. (4) Complete shift of immunodominance to epitope B, together with a complete replacement of the original variant in A by the escape mutant. Specific CTL to the original variant in epitope A, the escape variant in epitope A, and epitope B are denoted by x_1 , x_2 and y_1 , respectively.

dominance to epitope B. This happens if $1/c_1 + 1/c_2 > 1/k_1$ and $r_{11} > r_{21}$. The system converges to oscillations around an equilibrium with $v_{11}, v_{21}, x_1, y_1 > 0$ and $x_2 = 0$. (3) The new mutant may induce a new specific response to epitope A, which outcompetes the response against the original virus, and induce a partial shift in immunodominance. The conditions for this behaviour are $1/c_1 + 1/c_2 > 1/k_1 > 1/c_2$ and $r_{11} < r_{21}$. The system converges to oscillations around an equilibrium with $v_{11}, v_{21}, x_2, y_1 > 0$ and $x_1 = 0$. (4) Finally, the new mutant can induce a complete shift in immunodominance to epitope B. This happens for $1/c_2 > 1/k_1$ and $r_{11} < r_{21}$, and brings us eventually to oscillations around an equilibrium with $v_{21}, y_1 > 0$ and $v_{11}, x_1, x_2 = 0$. For cases (1) and (3), a homogeneous population of the mutant, v_{21} , would induce an immunodominant response against epitope A, whereas for case (4) it would induce immunodominance of epitope B. For case (2) it is not specified. Thus shifting immunodominance does not simply reflect the immunogenicity of the escape mutant.

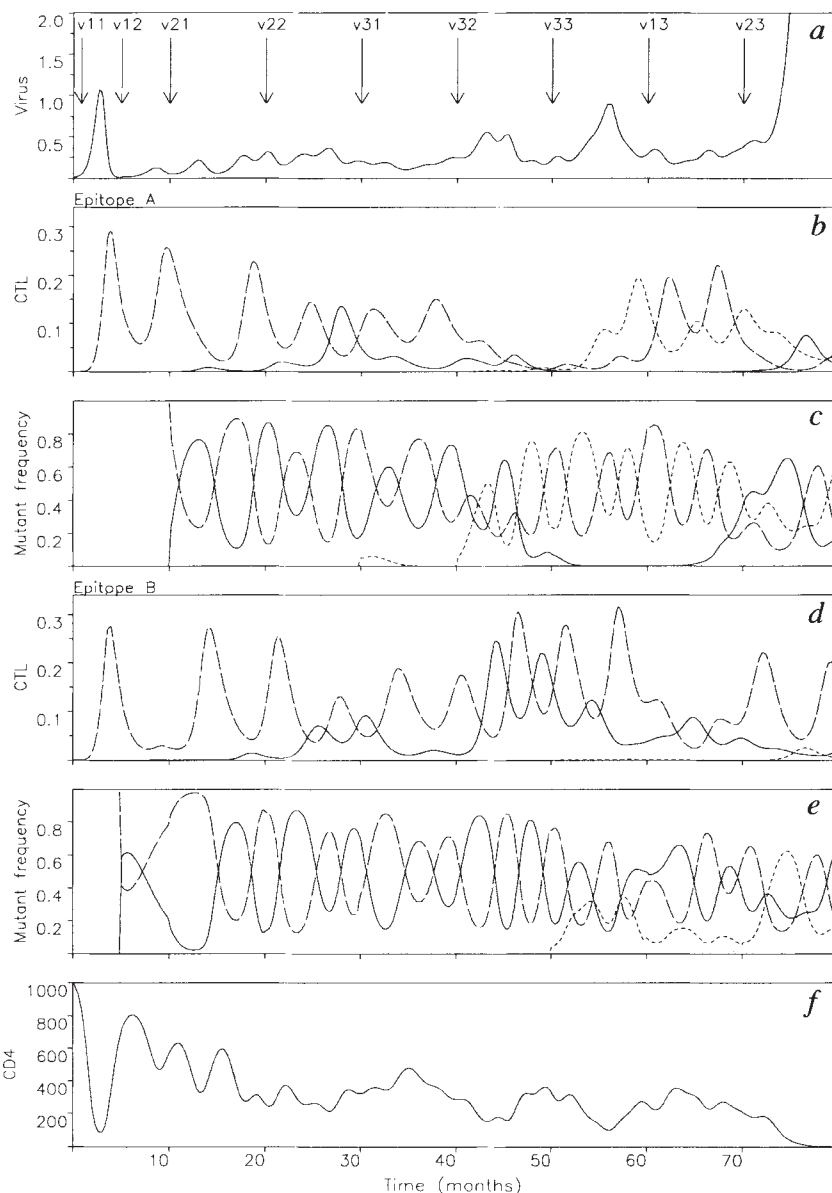
Patient 007 appears to represent an example of case (1). There are two variants in epitope p24.14, both of which are seen by the patient's CTL. There is no indication of a shifted response

to another epitope. Patient 020 provides examples of case (2). In all three epitopes there are variants that are not seen at some time points by the patient's CTL. There are shifts in the dominant response from one epitope to another.

It has often been asked how escape from CTL can be important if there are responses against several epitopes; our models show that persistent CTL responses against multiple epitopes are caused by immune escape. Furthermore, it is generally believed that patients who recognize a variety of different epitopes have better immune control than patients who only respond to a single epitope. Our theoretical and empirical studies suggest, however, that the opposite may be true, because persistent multiple-epitope responses are a consequence of antigenic variation.

The concept of shifting immunodominance in response to antigenic variation can also explain why not all escape mutants increase in abundance. They may remain at low levels, because the selection pressure on the virus population has changed with the altered immune response. Therefore, it is obvious that a conclusive demonstration of escape from immune response and selective increase in abundance is much more difficult than

FIG. 4 HIV disease progression can be modelled with two additional assumptions: (1) CTL proliferation requires $CD4^+$ help; and (2) HIV impairs $CD4^+$ cell function. This creates an upper limit of viral diversity beyond which the immune system can no longer control HIV replication. In this schematic simulation we consider CTL responses against two epitopes, A and B. With time we introduce escape mutants in both epitopes. This leads to a sequence of antigenic fluctuations and shifting immunodominance from one epitope to the other. Throughout most of the infection there are strong responses against both epitopes. The prevalence of individual virus mutants fluctuates and so does the specificity of the CTL response. Virus load increases with antigenic diversity. $CD4^+$ cells are declining in an almost linear fashion (subject to fluctuations that reflect fluctuations in viral load). The final mutant v_{23} breaches the diversity threshold, and the virus population escapes and destroys the last $CD4^+$ cells. The differential equations for this simulation are: $dv_{ij}/dt = v_{ij}(r_{ij} - px_i - qy_j)$, $dx_i/dt = \eta czv_{i*} + x_i(czv_{i*} - b)$, $dy_j/dt = \eta kzv_{j*} + y_j(kzv_{j*} - b)$, and $dz/dt = \lambda - dz - uvz$. The variable z denotes $CD4^+$ cell population size. For simplicity we assume that $CD4^+$ cells immigrate from precursor cells at a constant rate, λ , die at rate dz , and are killed by HIV at rate uvz . These population dynamics simply reflect the assumption that $CD4^+$ cell numbers and function are inversely correlated with virus load. All other variables and parameters are defined as in equations (1–3). a, Total virus abundance; b, CTL response against the variants of epitope A; c, frequency of the variants of epitope A; d, CTL response against the variants of epitope B; e, frequency of the variants of epitope B; f, $CD4^+$ T-cell count. We used the parameters: $p=1$, $q=1$, $c=2.98$, $k=2.96$, $b=0.1$, $\lambda=0.1$, $d=0.1$, $u=1$, $\eta=0.001$; the r_{ij} were chosen randomly between 0 and 0.2.



escape from drug treatment. Escape from drug treatment is the consequence of a constant selection pressure on the virus population, whereas escape from immune responses alters the selection pressure exerted by the immune system.

The model provides a theory for immunodominance and antigenic variation for variable pathogens. It is the first theoretical formulation of the escape dynamics of a pathogen from immune responses directed at multiple epitopes. Although the ideas have been developed in the specific context of HIV, our intention is to shed light on the general problem of immunodominance and antigenic variation with implications for broader classes of host-parasite systems.

With respect to HIV, we have shown that it is possible to explain much of the observed complexity of viral escape from CTL responses by a simpler nonlinear model. The model suggests that those patients who recognize fewer epitopes of the virus have a more stable CTL response and thus control the virus more effectively. This is consistent with the contrasting data for, and clinical history of, patients 007 and 020. It remains to be determined whether these patients are typical of slow or more rapid progressors towards AIDS. The model makes the testable prediction that a stable CTL response to an invariant epitope may lead to slow progression. Conversely, viral diversity and evolution during infection can lead to AIDS^{27,28}; the shift in CTL response towards weaker epitopes should increase overall viral loads and hence drive disease progression (Fig. 4). One testable route to immunotherapy would be to boost the CTL response to single conserved epitopes, making them immunodominant by increasing the frequency of reacting T cells,

thereby aiming to induce a stable pattern of immunological recognition. □

Received 19 January; accepted 4 May 1995.

- Hill, A. B. et al. *Immun. Rev.* **133**, 75–91 (1993).
- Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601–633 (1989).
- Adorini, L. et al. *J. exp. Med.* **168**, 2091–2104 (1988).
- Liu, Z. et al. *J. Immun.* **151**, 1852–1858 (1993).
- Sercarz, E. E. et al. *A. Rev. Immun.* **11**, 729–766 (1993).
- Buus, S. et al. *Science* **235**, 1353–1358 (1987).
- Schaeffer, E. B. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4649–4653 (1989).
- Zinkernagel, R. M. et al. *J. exp. Med.* **148**, 592–606 (1978).
- Takahashi, H. et al. *Science* **246**, 118–121 (1989).
- Gegin, C. & Lehmann-Grube, F. *J. Immun.* **149**, 3331–3338 (1992).
- Phillips, R. E. et al. *Nature* **354**, 453–459 (1991).
- Aebischer, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 11047–11051 (1991).
- Klenerman, P. et al. *Nature* **369**, 403–407 (1994).
- Niewiesk, S. et al. *J. Virol.* **69**, 2649–2653 (1995).
- Campos-Lima, P. et al. *Science* **260**, 98–100 (1993).
- Bertoletti, A. et al. *Nature* **369**, 407–410 (1994).
- Moskophidids, D. & Zinkernagel, R. M. *J. Virol.* **69**, 2187–2193 (1995).
- Nixon, D. F. et al. *Nature* **336**, 484–487 (1988).
- Walker, B. D. *Science* **240**, 64–66 (1988).
- McMichael, A. J. & Walker, B. D. *AIDS* **8**, S155–S174 (1994).
- Koup, R. A. et al. *J. Virol.* **68**, 4650–4655 (1994).
- Safrit, J. T. et al. *J. exp. Med.* **179**, 463–472 (1994).
- Pantaleo, G. et al. *Nature* **370**, 463–467 (1994).
- Nowak, M. A., May, R. M., Sigmund, K. J. *theor. Biol.* (in the press).
- Bangham, C. R. M. & McMichael, A. J. in *T-cell Immunity to Viruses in T Cells* (eds Feldmann, M., Lamb, J., Owen, M. J.) 281–310 (Wiley, New York, 1989).
- Carpenter, S. in *Applied Virology Research 2. Virus Variability, Epidemiology, and control* (eds Kurstak, E., Marusyk, R. G., Murphy, F. A. & van Regenmortel, H. V.) 99–115 (Plenum Medical Book Company, New York and London, 1990).
- Nowak, M. A. et al. *Science* **254**, 963–969 (1991).
- Nowak, M. A. & May, R. M. *AIDS* **7** (suppl. 1), S3–S18 (1993).

ACKNOWLEDGEMENTS. Support from the Wellcome Trust (M.A.N., R.E.P., D.G.L., P.K.), the MRC (A.J.M., S.R.J., S.M.), the Royal Society (R.M.M.) and Keble College (M.A.N.) is gratefully acknowledged. We thank P. Giangrande and all the staff and patients of the Haemophilia Centre, Churchill Hospital, Oxford, for their cooperation, and R. Anderson for discussion.

A DNA metalloenzyme with DNA ligase activity

Bernard Cuenoud* & Jack W. Szostak

Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

SINGLE-STRANDED DNA can fold into well-defined sequence-dependent tertiary structures that specifically bind a variety of target molecules^{1–10}, raising the possibility that some folded single-stranded DNAs might exhibit catalytic activities similar to those of ribozymes and protein enzymes. Derivatives of the hammerhead ribozyme that contain a majority of deoxyribonucleotides retain the ability to cleave RNA¹¹, and a 'deoxyribozyme' was generated by leaving all essential ribonucleotides of the hammerhead on the RNA 'substrate'¹². Recently *in vitro* selection has been used to isolate a DNA sequence that shows Pb²⁺-dependent RNA-cleaving activity¹³. Here we report the isolation by *in vitro* selection^{14–17} of a small single-stranded DNA that is a Zn²⁺/Cu²⁺-dependent metalloenzyme. The enzyme catalyses the formation of a new phosphodiester bond by the condensation of the 5'-hydroxyl of one oligodeoxynucleotide and a 3'-phosphorimidazole on another oligodeoxynucleotide, and shows multiple turnover ligation.

Oligodeoxynucleotides can be non-enzymatically ligated on either single-stranded¹⁸ or duplex¹⁹ DNA templates. We were interested in determining whether some DNA sequences might catalyse DNA ligation more efficiently than simple templating. We therefore designed an *in vitro* selection strategy for the isolation of such sequences from a large pool of random sequences

(Fig. 1), based on their ability to condense their own 5'-hydroxyl group to the chemically activated 3'-phosphate group of a substrate oligodeoxynucleotide. After nine cycles of selection and amplification, the DNA pool displayed efficient ligation activity (Fig. 2a). Incubation of pool 9 DNA with the activated substrate yielded a ligated product with the correct molecular mass and the expected sequence at the ligation junction. To analyse the selected sequences, DNA from pool 9 was cloned and sequenced. Most of the clones contained a common consensus sequence, consisting of two small domains separated by a spacer of variable length and sequence (Fig. 2b). These domains were embedded in entirely different flanking sequences, implying that several independent sequences in the original pool were carried through the selection process. Inspection of the consensus sequence suggests a secondary structure that is more complex than a simple template but nevertheless brings the 5'-hydroxyl and the 3'-phosphorimidazole into close proximity (Fig. 3a).

Based on the consensus sequence, we designed the small catalyst E47, a 47-nucleotide single-stranded DNA that ligates two separate DNA substrates S1 and S2 (Fig. 3b). Incubation of radioactively labelled S2 with activated substrate S1 and the catalyst E47 resulted in the appearance of the expected ligated product (Fig. 3c). Product formation required all three components to be present in the reaction, and required activation of the 3'-phosphate group of S1. E47 catalysed the ligation reaction twice as fast as pool 9. Small deletions within E47 resulted in severe losses of catalytic efficiency (Fig. 3d), indicating that the central consensus sequence is necessary for catalysis. The initial rate of ligation of S1 and S2 by E47 was 3,400-fold greater than the rate of the same reaction catalysed by a single complementary template under the same conditions, and was at least 10⁵-fold faster than the untemplated background ligation (Fig. 3d). This rate enhancement is comparable with values obtained for *in vitro* selected ribozymes^{14–17} and catalytic antibodies²⁰.

* Present address: Central Research Laboratories, Ciba-Geigy Limited, CH-4002 Basel, Switzerland.

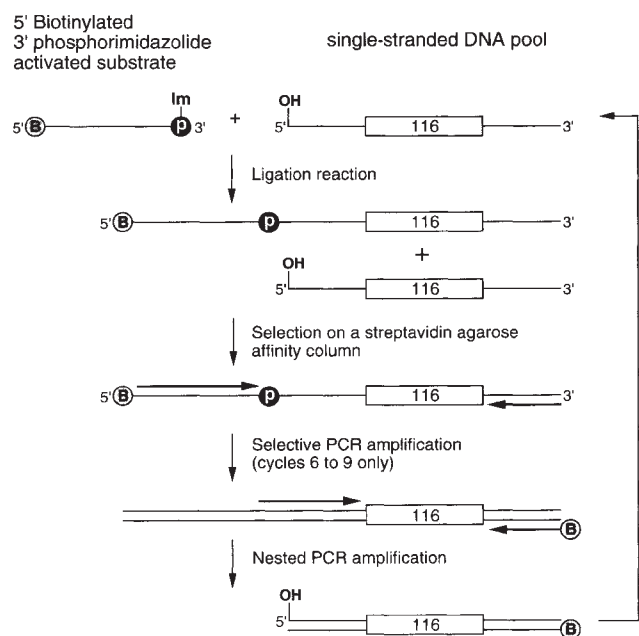


FIG. 1 Isolation of DNA sequences with ligase activity. The single-stranded (ss) DNA pool contained 116 random bases flanked by two constant regions complementary to the polymerase chain reaction (PCR) primers 5'-GGAACACTATCCGACTGGCACC-3' and 5'-biotin-CGGGATCC-TAATGACCAAGG-3'. The pool was prepared by solid-phase phosphoramidite chemistry and was PCR amplified⁴ to yield ~32 copies of 3.5×10^{14} different molecules. Single-stranded DNA was prepared as described³. The activated substrate bears a 5'-biotin group for selection purposes and a 3'-phosphorimidazolide²⁸ (5'-biotin-AAGCATCTAAGCAT-CTCAAGCplm). Eight copies of the DNA pool (0.5 μ M) were incubated with 1 μ M activated substrate and 1 μ M external template (5'-CGGATA-GTGTTCGGCTTGAGATGCTT-3', complementary to the 5' end of the pool and the 3' end of the substrate) in selection buffer (30 mM HEPES, pH 7.4, 600 mM KCl, 50 mM MgCl₂, 1 mM ZnCl₂). After 2 h the reaction was stopped by addition of EDTA. After 24 h, 0.5% ligated product was present. No product formation was observed in the absence of the external template. At cycle 7, pool activity was found to be independent of the external template, indicating that the remaining pool molecules were using an internal substrate binding site. In cycles 8 and 9, no external template was added, and reaction time was decreased to 2 and 0.5 min respectively to increase selection stringency. To isolate ligated molecules, reacted pool was passed through a streptavidin agarose affinity column (Pierce), unligated pool was washed off under denaturing conditions (3M urea followed by 150 mM NaOH, 40 column volumes each) and the ligated pool was specifically eluted with excess free biotin²⁹. To select for substrate ligation to the pool 5'-hydroxyl group, isolated DNA was selectively PCR amplified (cycles 6–9 only) with a primer with the substrate sequence and a second primer complementary to the 3' constant region, and gel purified. This pool was then subjected to nested PCR with the first set of primers, gel purified, and then re-amplified for ssDNA isolation³. Nine cycles of selection–amplification were performed, at which point the pool activity remained constant.

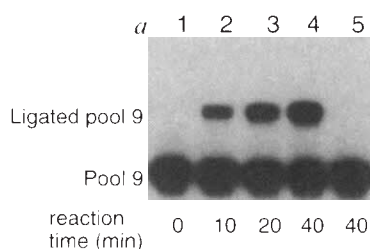
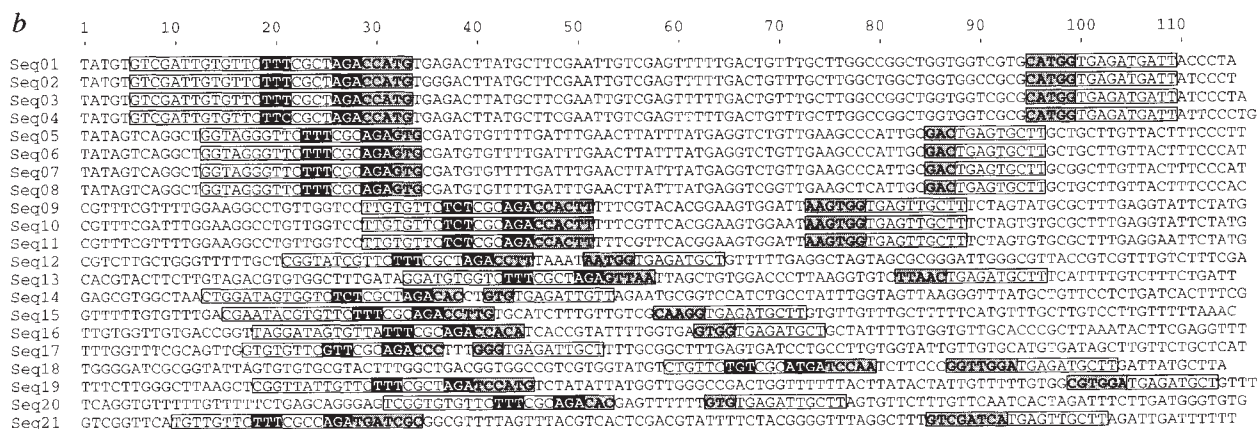


FIG. 2 a, Time course of reaction catalysed by pool 9 ssDNA. Internally labelled pool 9 DNA (0.5 μ M) was incubated with activated substrate (1 μ M) in selection buffer for the indicated time. Lane 5, substrate not activated. DNAs were separated by electrophoresis in a 6% polyacrylamide gel containing 8 M urea. Radioactivity detected by Molecular Dynamics Phosphorimager. b, DNA from pool 9 was PCR amplified and cloned into pT7Blue T-Vector (Novagen). Each clone was sequenced in both directions by the dideoxy method. The 21 sequences shown here (out of 26 in total) share a consensus sequence consisting of two conserved domains. Capital and lower-case letters in the consensus indicate highly and moderately conserved positions. X and Z define non-conserved but complementary bases. Boldface T in domain I is present in 50% of the clones, and absent in the others.



Consensus: 5'-(5-58)-cggataGTGTTCTTTCGCTAGACXXX- (2-60)-zzzzzTGAGATgctt- (4-62)-3'

Domain I Domain II

Because the catalyst is not consumed in the reaction, we expected E47 to be capable of catalysing the ligation of several molar equivalents of substrates S1 and S2, if the ligated product can dissociate from the enzyme. At saturating concentrations (140 μM) of both substrates and 1 μM E47, we observed multiple turnover catalysis at the rate of 0.66 h^{-1} at 25°C and 2.4 h^{-1} at 35°C (10 turnovers observed). At these temperatures, product release appears to be rate-limiting because we observed a rapid initial burst of approximately one equivalent of product formation within the first 10 minutes of the reaction. The initial rate of ligation in this burst phase was directly proportional to the concentration of E47 over a 30-fold range as expected for an enzyme at saturating substrate concentration²¹; a plot of V_{max} against E47 concentration yielded a k_{cat} of 4.2 h^{-1} (0.07 min^{-1}) at 25°C .

Because divalent metal ions play a crucial role in ribozymes²² and many protein enzymes²³, we expected the DNA catalyst to require Mg^{2+} and/or Zn^{2+} for activity, as these ions were present in the selection buffer. Indeed, the ligation reaction was dependent on Zn^{2+} (Fig. 4a) but did not require Mg^{2+} . We tested all the members of the Irving-Williams series (Ba^{2+} , Sr^{2+} , Ca^{2+} , Mg^{2+} , Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) as well as Pb^{2+} and Cd^{2+} between 10 μM and 10 mM, and found that only Cu^{2+}

could substitute for Zn^{2+} . The efficiency of the ligation reaction was highly dependent on the divalent metal ion concentration (Fig. 4b). Increasing concentrations of Zn^{2+} up to 4 mM enhanced activity, but at higher concentrations the activity dropped sharply, suggesting the existence of inhibitory metal binding sites. A similar concentration dependence was observed for copper, but at a 400-fold lower concentration. The metal ion specificity suggests the existence of one or more metal-ion binding sites with stringent geometrical and/or size requirements.

To gain insight into the ligation mechanism, we investigated the pH-rate profile of the reaction under pre-steady-state (single turnover) conditions (Fig. 4c). The bell-shaped profile displayed with Cu^{2+} suggests that the rate-limiting step of the ligation reaction depends in part on two ionizable groups, one acidic and one basic, raising the possibility of a general acid-base mechanism²¹ in which copper complexes are involved in proton transfer. Metal-ion hydroxides are thought to act as a general base in some ribozyme-mediated RNA cleavage reactions^{22,24,25}. However, we cannot rule out other possibilities such as pH-dependent folding effects²⁶.

These results show that single-stranded DNA can act as a metalloenzyme, without requiring a single ribonucleotide in

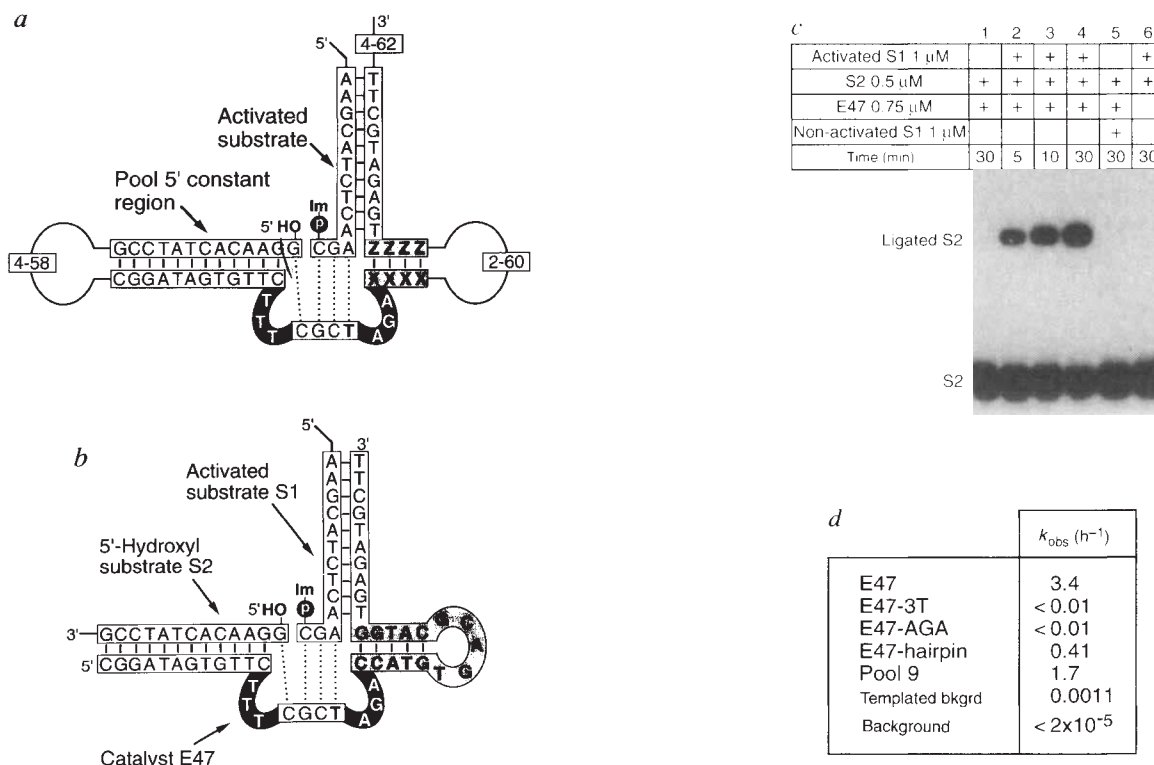


FIG. 3 **a**, Proposed secondary structure for the consensus sequence. The 5' end of domain I and the 3' end of domain II can form base pairs with the 5' constant region of the pool and the activated substrate respectively. The two complementary regions (grey boxes) form a stem and bring the flanking domains into close proximity. Dotted lines show possible interactions between the bases at the ligation junction and the sequence between the two boxed sequences TTT and AGA. **b**, Designed minimal DNA catalyst. Non-conserved regions in **a** were deleted to generate a three-fragment complex in which the formation of a phosphodiester bond between the 3'-phosphorimidazole substrate S1 and the 5'-hydroxyl substrate S2 is catalysed by the 47-nucleotide sequence E47. **c**, Time course of ligation of activated substrate S1 and radioactively labelled substrate S2 by the catalyst E47. No reaction was detectable when activated S1 or E47 were absent. **d**, Initial rates of ligation. The activated substrate S1 (1 μM) and radioactively labelled S2 (0.5 μM) were incubated with different catalysts (0.75 μM) at 25°C . The -3T and

-AGA constructs refer to deletion of TTT or AGA sequences from E47. In the -hairpin construct the hairpin has been replaced by 5'-CCATG. The background reaction with an external template (Fig. 1) was measured over a 6 h incubation. No product was detected in the absence of the template, corresponding to a maximum rate of $2 \times 10^{-5} \text{ h}^{-1}$.

METHODS. S2 was labelled at the 3'-end with [α - ^{32}P]-cordycepin-5'-triphosphate (NEN Dupont) and terminal transferase (Promega). Reaction conditions were as in Fig. 1 except that the selection buffer contained 30 mM HEPES, pH 7.2 and 4 mM ZnCl_2 . DNA was separated by electrophoresis on a 12% polyacrylamide gel containing 8 M urea. The k_{obs} values were determined by fitting fraction ligated against time to a linear equation by using KaleidaGraph, and are the average of two independent experiments measured at less than 20% product formation.

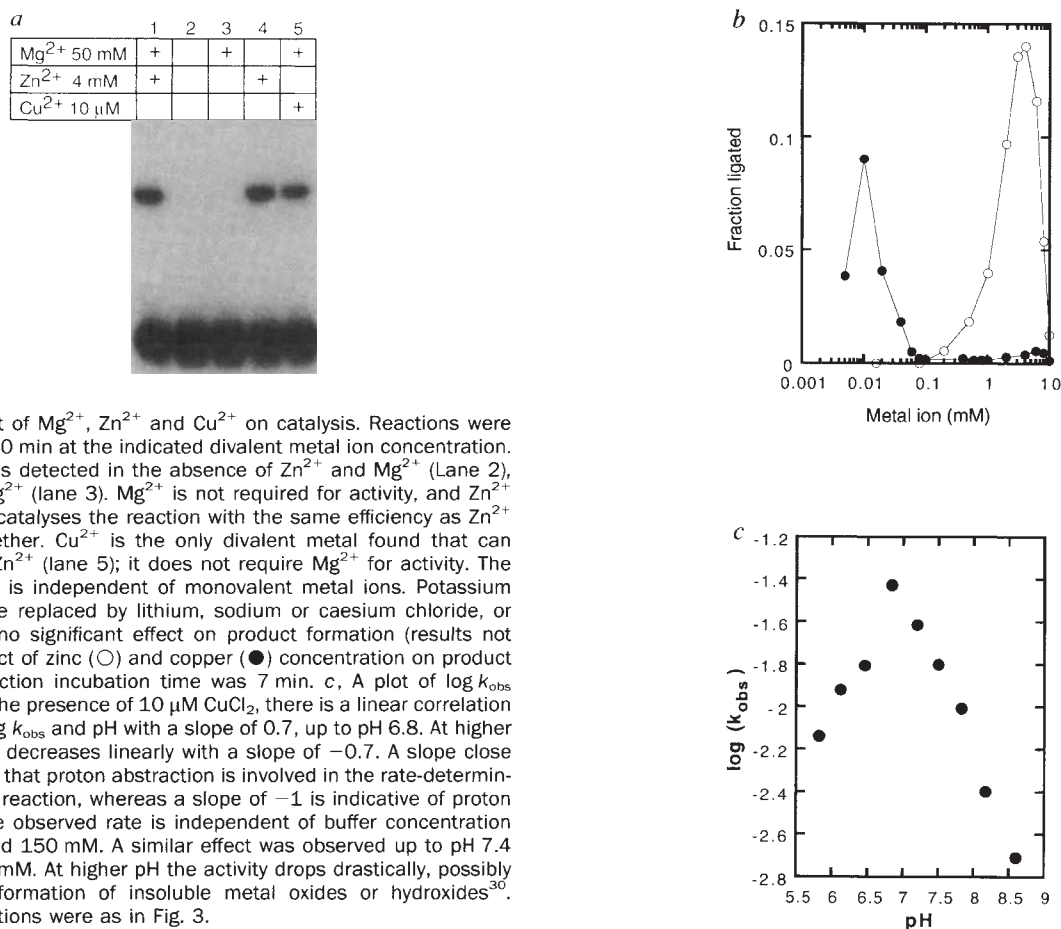


FIG. 4 *a*, Effect of Mg²⁺, Zn²⁺ and Cu²⁺ on catalysis. Reactions were incubated for 20 min at the indicated divalent metal ion concentration. No reaction was detected in the absence of Zn²⁺ and Mg²⁺ (Lane 2), or with only Mg²⁺ (lane 3). Mg²⁺ is not required for activity, and Zn²⁺ alone (lane 4) catalyses the reaction with the same efficiency as Zn²⁺ and Mg²⁺ together. Cu²⁺ is the only divalent metal found that can substitute for Zn²⁺ (lane 5); it does not require Mg²⁺ for activity. The rate of ligation is independent of monovalent metal ions. Potassium chloride can be replaced by lithium, sodium or caesium chloride, or removed with no significant effect on product formation (results not shown). *b*, Effect of zinc (○) and copper (●) concentration on product formation. Reaction incubation time was 7 min. *c*, A plot of log *k*_{obs} against pH. In the presence of 10 μM CuCl₂, there is a linear correlation between the log *k*_{obs} and pH with a slope of 0.7, up to pH 6.8. At higher pH, the activity decreases linearly with a slope of -0.7. A slope close to +1 suggests that proton abstraction is involved in the rate-determining step of the reaction, whereas a slope of -1 is indicative of proton donation²¹. The observed rate is independent of buffer concentration between 30 and 150 mM. A similar effect was observed up to pH 7.4 with Zn²⁺ at 4 mM. At higher pH the activity drops drastically, possibly owing to the formation of insoluble metal oxides or hydroxides³⁰. Reaction conditions were as in Fig. 3.

either the enzyme or the substrate for catalytic activity. *In vitro* selection techniques, along with the use of metal ions, cofactors and/or modified bases and backbones, should enable the isolation of DNA enzymes with diverse activities. If *in vitro* evolution can then lead to the identification of sequences with sufficient catalytic activity, the ease with which DNA oligonucleotides can be synthesized and their high stability could represent major advantages over other biopolymer catalysts for industrial and therapeutic applications.

Current theories of the origin of life emphasize the primordial role of RNA because of its unique ability both to catalyse reactions and to store information²⁷. However, it is now clear that DNA, and probably other related polynucleotides, can also fulfil this dual requirement. In view of the relative stability of DNA, it now seems important to investigate potential prebiotic pathways for the synthesis of deoxyribonucleotides, because even an inefficient pathway could allow the accumulation over time of significant concentrations of oligodeoxynucleotides. □

Received 8 February; accepted 13 April 1995.

1. Khan, A. S. & Roe, B. A. *Science* **241**, 74–76 (1988).
2. Perreault, J.-P. *et al. Eur. J. Biochem.* **186**, 87–93 (1989).
3. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. *Nature* **355**, 564–566 (1992).
4. Ellington, A. D. & Szostak, J. W. *Nature* **355**, 850–852 (1992).
5. Blackburn, E. H. A. *Rev. Biochem.* **61**, 113–129 (1992).

6. Wang, K. Y., Krawczyk, S. H., Bischofberger, N., Swaminathan, S. & Bolton, P. H. *Biochemistry* **32**, 11285–11292 (1993).
7. Paborsky, L. R., McCurdy, S. N., Griffin, L. C., Toole, J. J. & Leung, L. L. K. *J. biol. Chem.* **268**, 20808–20811 (1993).
8. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. *J. biol. Chem.* **268**, 17651–17654 (1993).
9. Wyatt, J. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 1356–1360 (1994).
10. Huizenga, D. E. & Szostak, J. W. *Biochemistry* **34**, 656–665 (1995).
11. Bratty, J., Chartrand, P., Ferbeyre, G. & Cedergren, R. *Biochim. biophys. Acta* **1216**, 345–359 (1993).
12. Chartrand, P., Harvey, S. C., Ferbeyre, G., Usman, N. & Cedergren, R. in *RNA Processing* 77 (Proc. RNA Soc., Madison, Wisconsin, 1994).
13. Breaker, R. R. & Joyce, G. F. *Chem. Biol.* **1**, 223–229 (1994).
14. Szostak, J. W. *Trends biochem. Sci.* **17**, 89–93 (1992).
15. Chapman, K. B. & Szostak, J. W. *Curr. Opin. Struct. Biol.* **4**, 618–622 (1994).
16. Breaker, R. R. & Joyce, G. F. *Trends Biotechnol.* **12**, 268–275 (1994).
17. Joyce, G. F. *Curr. Opin. struct. Biol.* **4**, 331–336 (1994).
18. Shabarova, Z. A. *Biochimie* **70**, 1323–1334 (1988).
19. Luecke, K. J. & Dervan, P. B. *J. Am. chem. Soc.* **111**, 8733–8735 (1989).
20. Lerner, R. A., Benkovic, S. J. & Schultz, P. G. *Science* **252**, 659–667 (1991).
21. Fersht, A. *Enzyme Structure and Mechanism* (Freeman, New York, 1985).
22. Pyle, A. M. *Science* **261**, 709–714 (1993).
23. Karlin, K. D. *Science* **261**, 701–708 (1993).
24. Dahm, S. C., Derrick, W. B. & Uhlenbeck, O. C. *Biochemistry* **32**, 13040–13045 (1993).
25. Pan, T., Ditch, B. & Uhlenbeck, O. C. *Biochemistry* **33**, 9561–9565 (1994).
26. Kao, T. H. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3360–3364 (1980).
27. Orgel, L. E. *Science* **271**(4), 76–83 (1994).
28. Chu, B. C. F., Kramer, F. R. & Orgel, L. E. *Nucleic Acids Res.* **14**, 5591–5603 (1986).
29. Wilson, C. & Szostak, J. W. *Nature* **374**, 777–782 (1995).
30. Bailar, J. C. Jr., Emeléus, H. J. & Trotman-Dickenson, A. F. *Comprehensive Inorganic Chemistry* (Pergamon, Elmsford, N.Y., 1973).

ACKNOWLEDGEMENTS. We thank members of our laboratory and A. Raymond for helpful suggestions and for critically reading the manuscript. B.C. was supported by a fellowship from the Swiss National Fund. This work was supported by grants from Hoechst AG and the NIH.